PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 :
G01N 33/536, C07K 14/00, C12N 15/00

A1

(11) International Publication Number: WO 00/34784

(43) International Publication Date: 15 June 2000 (15.06.00)

(21) International Application Number: PCT/US99/29317

(22) International Filing Date: 9 December 1999 (09.12.99)

(30) Priority Data:
60/111,737 10 December 1998 (10.12.98) US

(71) Applicant: PHYLOS, INC. [US/US]; 128 Spring Street, Lexington, MA 02421 (US).

(72) Inventor: LIPOVSEK, Dasa; 45 Sunset Road, Cambridge, MA 02138 (US).

(74) Agent: ELBING, Karen; Clark & Elbing LLP, 176 Federal Street, Boston, MA 02110–2214 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Published**

*With international search report.*

*Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.*

(54) Title: PROTEIN SCAFFOLDS FOR ANTIBODY MIMICS AND OTHER BINDING PROTEINS

(57) Abstract

Disclosed herein are proteins that include a fibronectin type III domain having at least one randomized loop. Also-disclosed herein are nucleic acids encoding such proteins and the use of such proteins in methods for evolving novel compound–binding species and their ligands.

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **AL** | Albania | **ES** | Spain | **LS** | Lesotho | **SI** | Slovenia |
| **AM** | Armenia | **FI** | Finland | **LT** | Lithuania | **SK** | Slovakia |
| **AT** | Austria | **FR** | France | **LU** | Luxembourg | **SN** | Senegal |
| **AU** | Australia | **GA** | Gabon | **LV** | Latvia | **SZ** | Swaziland |
| **AZ** | Azerbaijan | **GB** | United Kingdom | **MC** | Monaco | **TD** | Chad |
| **BA** | Bosnia and Herzegovina | **GE** | Georgia | **MD** | Republic of Moldova | **TG** | Togo |
| **BB** | Barbados | **GH** | Ghana | **MG** | Madagascar | **TJ** | Tajikistan |
| **BE** | Belgium | **GN** | Guinea | **MK** | The former Yugoslav Republic of Macedonia | **TM** | Turkmenistan |
| **BF** | Burkina Faso | **GR** | Greece | | | **TR** | Turkey |
| **BG** | Bulgaria | **HU** | Hungary | **ML** | Mali | **TT** | Trinidad and Tobago |
| **BJ** | Benin | **IE** | Ireland | **MN** | Mongolia | **UA** | Ukraine |
| **BR** | Brazil | **IL** | Israel | **MR** | Mauritania | **UG** | Uganda |
| **BY** | Belarus | **IS** | Iceland | **MW** | Malawi | **US** | United States of America |
| **CA** | Canada | **IT** | Italy | **MX** | Mexico | **UZ** | Uzbekistan |
| **CF** | Central African Republic | **JP** | Japan | **NE** | Niger | **VN** | Viet Nam |
| **CG** | Congo | **KE** | Kenya | **NL** | Netherlands | **YU** | Yugoslavia |
| **CH** | Switzerland | **KG** | Kyrgyzstan | **NO** | Norway | **ZW** | Zimbabwe |
| **CI** | Côte d'Ivoire | **KP** | Democratic People's Republic of Korea | **NZ** | New Zealand | | |
| **CM** | Cameroon | | | **PL** | Poland | | |
| **CN** | China | **KR** | Republic of Korea | **PT** | Portugal | | |
| **CU** | Cuba | **KZ** | Kazakstan | **RO** | Romania | | |
| **CZ** | Czech Republic | **LC** | Saint Lucia | **RU** | Russian Federation | | |
| **DE** | Germany | **LI** | Liechtenstein | **SD** | Sudan | | |
| **DK** | Denmark | **LK** | Sri Lanka | **SE** | Sweden | | |
| **EE** | Estonia | **LR** | Liberia | **SG** | Singapore | | |

# PROTEIN SCAFFOLDS FOR ANTIBODY MIMICS AND OTHER BINDING PROTEINS

## Background of the Invention

This invention relates to protein scaffolds useful, for example, for the generation of products having novel binding characteristics.

Proteins having relatively defined three-dimensional structures, commonly referred to as protein scaffolds, may be used as reagents for the design of engineered products. These scaffolds typically contain one or more regions which are amenable to specific or random sequence variation, and such sequence randomization is often carried out to produce libraries of proteins from which desired products may be selected. One particular area in which such scaffolds are useful is the field of antibody design.

A number of previous approaches to the manipulation of the mammalian immune system to obtain reagents or drugs have been attempted. These have included injecting animals with antigens of interest to obtain mixtures of polyclonal antibodies reactive against specific antigens, production of monoclonal antibodies in hybridoma cell culture (Koehler and Milstein, Nature 256:495, 1975), modification of existing monoclonal antibodies to obtain new or optimized recognition properties, creation of novel antibody fragments with desirable binding characteristics, and randomization of single chain antibodies (created by connecting the variable regions of the heavy and light chains of antibody molecules with a flexible peptide linker) followed by selection for antigen binding by phage display (Clackson et al., Nature 352:624, 1991).

In addition, several non-immunoglobulin protein scaffolds have been proposed for obtaining proteins with novel binding properties. For example, a

"minibody" scaffold, which is related to the immunoglobulin fold, has been designed by deleting three beta strands from a heavy chain variable domain of a monoclonal antibody (Tramontano et al., J. Mol. Recognit. 7:9, 1994). This protein includes 61 residues and can be used to present two hypervariable loops. These two loops have been randomized and products selected for antigen binding, but thus far the framework appears to have somewhat limited utility due to solubility problems. Another framework used to display loops has been tendamistat, a 74 residue, six-strand beta sheet sandwich held together by two disulfide bonds (McConnell and Hoess, J. Mol. Biol. 250:460, 1995). This scaffold includes three loops, but, to date, only two of these loops have been examined for randomization potential.

Other proteins have been tested as frameworks and have been used to display randomized residues on alpha helical surfaces (Nord et al., Nat. Biotechnol. 15:772, 1997; Nord et al., Protein Eng. 8:601, 1995), loops between alpha helices in alpha helix bundles (Ku and Schultz, Proc. Natl. Acad. Sci. USA 92:6552, 1995), and loops constrained by disulfide bridges, such as those of the small protease inhibitors (Markland et al., Biochemistry 35:8045, 1996; Markland et al., Biochemistry 35:8058, 1996; Rottgen and Collins, Gene 164:243, 1995; Wang et al., J. Biol. Chem. 270:12250, 1995).

## Summary of the Invention

The present invention provides a new family of proteins capable of evolving to bind any compound of interest. These proteins, which make use of a fibronectin or fibronectin-like scaffold, function in a manner characteristic of natural or engineered antibodies (that is, polyclonal, monoclonal, or single-chain antibodies) and, in addition, possess structural advantages. Specifically, the structure of these antibody mimics has been designed for optimal folding, stability, and solubility, even under conditions which normally

lead to the loss of structure and function in antibodies.

These antibody mimics may be utilized for the purpose of designing proteins which are capable of binding to virtually any compound (for example, any protein) of interest. In particular, the fibronectin-based molecules described herein may be used as scaffolds which are subjected to directed evolution designed to randomize one or more of the three fibronectin loops which are analogous to the complementarity-determining regions (CDRs) of an antibody variable region. Such a directed evolution approach results in the production of antibody-like molecules with high affinities for antigens of interest. In addition, the scaffolds described herein may be used to display defined exposed loops (for example, loops previously randomized and selected on the basis of antigen binding) in order to direct the evolution of molecules that bind to such introduced loops. A selection of this type may be carried out to identify recognition molecules for any individual CDR-like loop or, alternatively, for the recognition of two or all three CDR-like loops combined into a non-linear epitope.

Accordingly, the present invention features a protein that includes a fibronectin type III domain having at least one randomized loop, the protein being characterized by its ability to bind to a compound that is not bound by the corresponding naturally-occurring fibronectin.

In preferred embodiments, the fibronectin type III domain is a mammalian (for example, a human) fibronectin type III domain; and the protein includes the tenth module of the fibronectin type III ($^{10}$Fn3) domain. In such proteins, compound binding is preferably mediated by either one, two, or three $^{10}$Fn3 loops. In other preferred embodiments, the second loop of $^{10}$Fn3 may be extended in length relative to the naturally-occurring module, or the $^{10}$Fn3 may lack an integrin-binding motif. In these molecules, the integrin-binding motif may be replaced by an amino acid sequence in which a basic amino acid-

neutral amino acid-acidic amino acid sequence (in the N-terminal to C-terminal direction) replaces the integrin-binding motif; one preferred sequence is serine-glycine-glutamate. In another preferred embodiment, the fibronectin type III domain-containing proteins of the invention lack disulfide bonds.

Any of the fibronectin type II domain-containing proteins described herein may be formulated as part of a fusion protein (for example, a fusion protein which further includes an immunoglobulin $F_c$ domain, a complement protein, a toxin protein, or an albumin protein). In addition, any of the fibronectin type III domain proteins may be covalently bound to a nucleic acid (for example, an RNA), and the nucleic acid may encode the protein. Moreover, the protein may be a multimer, or, particularly if it lacks an integrin-binding motif, it may be formulated in a physiologically-acceptable carrier.

The present invention also includes features proteins that include a fibronectin type III domain having at least one mutation in a β-sheet sequence which changes the scaffold structure. Again, these proteins are characterized by their ability to bind to compound that are not bound by the corresponding naturally-occurring fibronectin.

In a related aspect, the invention further features nucleic acids encoding any of the proteins of the invention. In preferred embodiments, the nucleic acid is DNA or RNA.

In another related aspect, the invention also features a method for generating a protein which includes a fibronectin type III domain and which is pharmaceutically acceptable to a mammal, involving removing the integrin-binding domain of said fibronectin type III domain. This method may be applied to any of the fibronectin type III domain-containing proteins described above and is particularly useful for generating proteins for human therapeutic applications. The invention also features such fibronectin type III domain-containing proteins which lack integrin-binding domains.

In yet other related aspects, the invention features screening methods which may be used to obtain or evolve randomized fibronectin type III proteins capable of binding to compounds of interest, or to obtain or evolve compounds (for example, proteins) capable of binding to a particular protein containing a randomized fibronectin type III motif. In addition, the invention features screening procedures which combine these two methods, in any order, to obtain either compounds or proteins of interest.

In particular, the first screening method, useful for the isolation or identification of randomized proteins of interest, involves : (a) contacting the compound with a candidate protein, the candidate protein including a fibronectin type III domain having at least one randomized loop, the contacting being carried out under conditions that allow compound-protein complex formation; and (b) obtaining, from the complex, the protein which binds to the compound.

The second screening method, for isolating or identifying a compound which binds to a protein having a randomized fibronectin type III domain, involves: (a) contacting the protein with a candidate compound, the contacting being carried out under conditions that allow compound-protein complex formation; and (b) obtaining, from the complex, the compound which binds to the protein.

In preferred embodiments, the methods further involve either randomizing at least one loop of the fibronectin type III domain of the protein obtained in step (b) and repeating steps (a) and (b) using the further randomized protein, or modifying the compound obtained in step (b) and repeating steps (a) and (b) using the further modified compound. In addition, the compound is preferably a protein, and the fibronectin type III domain is preferably a mammalian (for example, a human) fibronectin type III domain. In other preferred embodiments, the protein includes the tenth module of the fibronectin

type III domain ($^{10}$Fn3), and binding is mediated by one, two or three $^{10}$Fn3 loops. In addition, the second loop of $^{10}$Fn3 may be extended in length relative to the naturally-occurring module, or $^{10}$Fn3 may lack an integrin-binding motif. Again, as described above, the integrin-binding motif may be replaced by an amino acid sequence in which a basic amino acid-neutral amino acid-acidic amino acid sequence (in the N-terminal to C-terminal direction) replaces the integrin-binding motif; one preferred sequence is serine-glycine-glutamate.

The selection methods described herein may be carried out using any fibronectin type III domain-containing protein. For example, the fibronectin type III domain-containing protein may lack disulfide bonds, or may be formulated as part of a fusion protein (for example, a fusion protein which further includes an immunoglobulin $F_c$ domain, a complement protein, a toxin protein, or an albumin protein). In addition, selections may be carried out using the fibronectin type III domain proteins covalently bound to nucleic acids (for example, RNAs or any nucleic acid which encodes the protein). Moreover, the selections may be carried out using fibronectin domain-containing protein multimers.

Preferably, the selections involve the immobilization of the binding target on a solid support. Preferred solid supports include columns (for example, affinity columns, such as agarose columns) or microchips.

As used herein, by "fibronectin type III domain" is meant a domain having 7 or 8 beta strands which are distributed between two beta sheets, which themselves pack against each other to form the core of the protein, and further containing loops which connect the beta strands to each other and are solvent exposed. There are at least three such loops at each edge of the beta sheet sandwich, where the edge is the boundary of the protein perpendicular to the direction of the beta strands. Preferably, a fibronectin type III domain includes a sequence which exhibits at least 30% amino acid identity, and preferably at

least 50% amino acid identity, to the sequence encoding the structure of the $^{10}$Fn3 domain referred to as "1ttg" (ID = "1ttg" (one ttg)) available from the Protein Data Base. Sequence identity referred to in this definition is determined by the Homology program, available from Molecular Simulation (San Diego, CA). The invention further includes polymers of $^{10}$Fn3-related molecules, which are an extension of the use of the monomer structure, whether or not the subunits of the polyprotein are identical or different in sequence.

By "naturally occurring fibronectin" is meant any fibronectin protein that is encoded by a living organism.

By "randomized" is meant including one or more amino acid alterations relative to a template sequence.

By a "protein" is meant any sequence of two or more amino acids, regardless of length, post-translation modification, or function. "Protein" and "peptide" are used interchangeably herein.

By "RNA" is meant a sequence of two or more covalently bonded, naturally occurring or modified ribonucleotides. One example of a modified RNA included within this term is phosphorothioate RNA.

By "DNA" is meant a sequence of two or more covalently bonded, naturally occurring or modified deoxyribonucleotides.

By a "nucleic acid" is meant any two or more covalently bonded nucleotides or nucleotide analogs or derivatives. As used herein, this term includes, without limitation, DNA, RNA, and PNA.

By "pharmaceutically acceptable" is meant a compound or protein that may be administered to an animal (for example, a mammal) without significant adverse medical consequences.

By "physiologically acceptable carrier" is meant a carrier which does not have a significant detrimental impact on the treated host and which retains the therapeutic properties of the compound with which it is administered. One

exemplary physiologically acceptable carrier is physiological saline. Other physiologically acceptable carriers and their formulations are known to one skilled in the art and are described, for example, in Remington's Pharmaceutical Sciences, (18th edition), ed. A. Gennaro, 1990, Mack Publishing Company, Easton, PA, incorporated herein by reference.

By "selecting" is meant substantially partitioning a molecule from other molecules in a population. As used herein, a "selecting" step provides at least a 2-fold, preferably, a 30-fold, more preferably, a 100-fold, and, most preferably, a 1000-fold enrichment of a desired molecule relative to undesired molecules in a population following the selection step. A selection step may be repeated any number of times, and different types of selection steps may be combined in a given approach.

By "binding partner," as used herein, is meant any molecule which has a specific, covalent or non-covalent affinity for a portion of a desired compound (for example, protein) of interest. Examples of binding partners include, without limitation, members of antigen/antibody pairs, protein/inhibitor pairs, receptor/ligand pairs (for example cell surface receptor/ligand pairs, such as hormone receptor/peptide hormone pairs), enzyme/substrate pairs (for example, kinase/substrate pairs), lectin/carbohydrate pairs, oligomeric or heterooligomeric protein aggregates, DNA binding protein/DNA binding site pairs, RNA/protein pairs, and nucleic acid duplexes, heteroduplexes, or ligated strands, as well as any molecule which is capable of forming one or more covalent or non-covalent bonds (for example, disulfide bonds) with any portion of another molecule (for example, a compound or protein).

By a "solid support" is meant, without limitation, any column (or column material), bead, test tube, microtiter dish, solid particle (for example, agarose or sepharose), microchip (for example, silicon, silicon-glass, or gold

chip), or membrane (for example, the membrane of a liposome or vesicle) to which an affinity complex may be bound, either directly or indirectly (for example, through other binding partner intermediates such as other antibodies or Protein A), or in which an affinity complex may be embedded (for example, through a receptor or channel).

The present invention provides a number of advantages. For example, as described in more detail below, the present antibody mimics exhibit improved biophysical properties, such as stability under reducing conditions and solubility at high concentrations. In addition, these molecules may be readily expressed and folded in prokaryotic systems, such as E. coli, in eukaryotic systems, such as yeast, and in in vitro translation systems, such as the rabbit reticulocyte lysate system. Moreover, these molecules are extremely amenable to affinity maturation techniques involving multiple cycles of selection, including in vitro selection using RNA-protein fusion technology (Roberts and Szostak, Proc. Natl. Acad. Sci USA 94:12297, 1997; Szostak et al., U.S.S.N. 09/007,005 and U.S.S.N. 09/247,190; Szostak et al. WO98/31700), phage display (see, for example, Smith and Petrenko, Chem. Rev. 97:317, 1997), and yeast display systems (see, for example, Boder and Wittrup, Nature Biotech. 15:553, 1997).

Other features and advantages of the present invention will be apparent from the following detailed description thereof, and from the claims.

## Brief Description of the Drawings

FIGURE 1 is a photograph showing a comparison between the structures of antibody heavy chain variable regions from camel (dark blue) and llama (light blue), in each of two orientations.

FIGURE 2 is a photograph showing a comparison between the

structures of the camel antibody heavy chain variable region (dark blue), the llama antibody heavy chain variable region (light blue), and a fibronectin type III module number 10 ($^{10}$Fn3) (yellow).

FIGURE 3 is a photograph showing a fibronectin type III module number 10 ($^{10}$Fn3), with the loops corresponding to the antigen-binding loops in IgG heavy chains highlighted in red.

FIGURE 4 is a graph illustrating a sequence alignment between a fibronectin type III protein domain and related protein domains.

FIGURE 5 is a photograph showing the structural similarities between a $^{10}$Fn3 domain and 15 related proteins, including fibronectins, tenascins, collagens, and undulin. In this photograph, the regions are labeled as follows: constant, dark blue; conserved, light blue; neutral, white; variable, red; and RGB integrin-binding motif (variable), yellow.

FIGURE 6 is a photograph showing space filling models of fibronectin III modules 9 and 10, in each of two different orientations. The two modules and the integrin binding loop (RGB) are labeled. In this figure, blue indicates positively charged residues, red indicates negatively charged residues, and white indicates uncharged residues.

FIGURE 7 is a photograph showing space filling models of fibronectin III modules 7-10, in each of three different orientiations. The four modules are labeled. In this figure, blue indicates positively charged residues, red indicates negatively charged residues, and white indicates uncharged residues.

FIGURE 8 is a photograph illustrating the formation, under different salt conditions, of RNA-protein fusions which include fibronectin type III domains.

FIGURE 9 is a series of photographs illustrating the selection of fibronectin type III domain-containing RNA-protein fusions, as measured by

PCR signal analysis.

FIGURE 10 is a graph illustrating an increase in the percent TNF-α binding during the selections described herein, as well as a comparison between RNA-protein fusion and free protein selections.

FIGURE 11 is a series of schematic representations showing IgG, $^{10}$Fn3, Fn-$CH_1$-$CH_2$-$CH_3$, and Fn-$CH_2$-$CH_3$ (clockwise from top left).

FIGURE 12 is a photograph showing a molecular model of Fn-$CH_1$-$CH_2$-$CH_3$ based on known three-dimensional structures of IgG (X-ray crystallography) and $^{10}$Fn3 (NMR and X-ray crystallography).

## Detailed Description

The novel antibody mimics described herein have been designed to be superior both to antibody-derived fragments and to non-antibody frameworks, for example, those frameworks described above.

The major advantage of these antibody mimics over antibody fragments is structural. These scaffolds are derived from whole, stable, and soluble structural modules found in human body fluid proteins. Consequently, they exhibit better folding and thermostability properties than antibody fragments, whose creation involves the removal of parts of the antibody native fold, often exposing amino acid residues that, in an intact antibody, would be buried in a hydrophobic environment, such as an interface between variable and constant domains. Exposure of such hydrophobic residues to solvent increases the likelihood of aggregation.

In addition, the antibody mimics described herein have no disulfide bonds, which have been reported to retard or prevent proper folding of antibody fragments under certain conditions. Since the present scaffolds do not rely on disulfides for native fold stability, they are stable under reducing conditions, unlike antibodies and their fragments which unravel upon disulfide bond

breakdown.

Moreover, these fibronectin-based scaffolds provide the functional advantages of antibody molecules. In particular, despite the fact that the $^{10}$Fn3 module is not an immunoglobulin, its overall fold is close to that of the variable region of the IgG heavy chain (Figure 2), making it possible to display the three fibronectin loops analogous to CDRs in relative orientations similar to those of native antibodies. Because of this structure, the present antibody mimics possess antigen binding properties that are similar in nature and affinity to those of antibodies, and a loop randomization and shuffling strategy may be employed in vitro that is similar to the process of affinity maturation of antibodies in vivo.

There are now described below exemplary fibronectin-based scaffolds and their use for identifying, selecting, and evolving novel binding proteins as well as their target ligands. These examples are provided for the purpose of illustrating, and not limiting, the invention.

$^{10}$Fn3 Structural Motif

The antibody mimics of the present invention are based on the structure of a fibronectin module of type III (Fn3), a common domain found in mammalian blood and structural proteins. This domain occurs more than 400 times in the protein sequence database and has been estimated to occur in 2% of the proteins sequenced to date, including fibronectins, tenscin, intracellular cytoskeletal proteins, and prokaryotic enzymes (Bork and Doolittle, Proc. Natl. Acad. Sci. USA 89:8990, 1992; Bork et al., Nature Biotech. 15:553, 1997; Meinke et al., J. Bacteriol. 175:1910, 1993; Watanabe et al., J. Biol. Chem. 265:15659, 1990). In particular, these scaffolds include, as templates, the tenth module of human Fn3 ($^{10}$Fn3), which comprises 94 amino acid residues. The overall fold of this domain is closely related to that of the smallest functional

antibody fragment, the variable region of the heavy chain, which comprises the entire antigen recognition unit in camel and llama IgG (Figure 1, 2). The major differences between camel and llama domains and the $^{10}$Fn3 domain are that (i) $^{10}$Fn3 has fewer beta strands (seven vs. nine) and (ii) the two beta sheets packed against each other are connected by a disulfide bridge in the camel and llama domains, but not in $^{10}$Fn3.

The three loops of $^{10}$Fn3 corresponding to the antigen-binding loops of the IgG heavy chain run between amino acid residues 21-31, 51-56, and 76-88 (Figure 3). The length of the first and the third loop, 11 and 12 residues, respectively, fall within the range of the corresponding antigen-recognition loops found in antibody heavy chains, that is, 10-12 and 3-25 residues, respectively. Accordingly, once randomized and selected for high antigen affinity, these two loops make contacts with antigens equivalent to the contacts of the corresponding loops in antibodies.

In contrast, the second loop of $^{10}$Fn3 is only 6 residues long, whereas the corresponding loop in antibody heavy chains ranges from 16-19 residues. To optimize antigen binding, therefore, the second loop of $^{10}$Fn3 is preferably extended by 10-13 residues (in addition to being randomized) to obtain the greatest possible flexibility and affinity in antigen binding. Indeed, in general, the lengths as well as the sequences of the CDR-like loops of the antibody mimics may be randomized during in vitro or in vivo affinity maturation (as described in more detail below).

The tenth human fibronectin type III domain, $^{10}$Fn3, refolds rapidly even at low temperature; its backbone conformation has been recovered within 1 second at 5°C. Thermodynamic stability of $^{10}$Fn3 is high ($\Delta G_U$ = 24 kJ/mol = 5.7 kcal/mol), correlating with its high melting temperature of 110°C.

One of the physiological roles of $^{10}$Fn3 is as a subunit of fibronectin, a glycoprotein that exists in a soluble form in body fluids and in an insoluble

form in the extracellular matrix (Dickinson et al., J. Mol. Biol. 236:1079, 1994). A fibronectin monomer of 220-250 kD contains 12 type I modules, two type II modules, and 17 fibronectin type III modules (Potts and Campbell, Curr. Opin.Cell Biol. 6:648, 1994). Different type III modules are involved in the binding of fibronectin to integrins, heparin, and chondroitin sulfate. $^{10}$Fn3 was found to mediate cell adhesion through an integrin-binding Arg-Gly-Asp (RGD) motif on one of its exposed loops. Similar RGD motifs have been shown to be involved in integrin binding by other proteins, such as fibrinogen, von Wellebrand factor, and vitronectin (Hynes et al., Cell 69:11, 1992). No other matrix- or cell-binding roles have been described for $^{10}$Fn3.

The observation that $^{10}$Fn3 has only slightly more adhesive activity than a short peptide containing RGD is consistent with the conclusion that the cell-binding activity of $^{10}$Fn3 is localized in the RGD peptide rather than distributed throughout the $^{10}$Fn3 structure (Baron et al., Biochemistry 31:2068, 1992). The fact that $^{10}$Fn3 without the RGD motif is unlikely to bind to other plasma proteins or extracellular matrix makes $^{10}$Fn3 a useful scaffold to replace antibodies. In addition, the presence of $^{10}$Fn3 in natural fibrinogen in the bloodstream suggests that $^{10}$Fn3 itself is unlikely to be immunogenic in the organism of origin.

In addition, we have determined that the $^{10}$Fn3 framework possesses exposed loop sequences tolerant of randomization, facilitating the generation of diverse pools of antibody mimics. This determination was made by examining the flexibility of the $^{10}$Fn3 sequence. In particular, the human $^{10}$Fn3 sequence was aligned with the sequences of fibronectins from other sources as well as sequences of related proteins (Figure 4), and the results of this alignment were mapped onto the three-dimensional structure of the human $^{10}$Fn3 domain (Figure 5). This alignment revealed that the majority of conserved residues are found in the core of the beta sheet sandwich, whereas the highly variable

residues are located along the edges of the beta sheets, including the N- and C-termini, on the solvent-accessible faces of both beta sheets, and on three solvent-accessible loops that serve as the hypervariable loops for affinity maturation of the antibody mimics. In view of these results, the randomization of these three loops are unlikely to have an adverse effect on the overall fold or stability of the $^{10}$Fn3 framework itself.

For the human $^{10}$Fn3 sequence, this analysis indicates that, at a minimum, amino acids 1-9, 44-50, 61-54, 82-94 (edges of beta sheets); 19, 21, 30-46 (even), 79-65 (odd) (solvent-accessible faces of both beta sheets); 21-31, 51-56, 76-88 (CDR-like solvent-accessible loops); and 14-16 and 36-45 (other solvent-accessible loops and beta turns) may be randomized to evolve new or improved compound-binding proteins. In addition, as discussed above, alterations in the lengths of one or more solvent exposed loops may also be included in such directed evolution methods. Alternatively, changes in the β-sheet sequences may also be used to evolve new proteins. These mutations change the scaffold and thereby indirectly alter loop structure(s). If this approach is taken, mutations should not saturate the sequence, but rather few mutations should be introduced. Preferably, no more than 10 amino acid changes, and, more preferably, no more than 3 amino acid changes should be introduced to the β-sheet sequences by this approach.

Fibronectin Fusions

The antibody mimics described herein may be fused to other protein domains. For example, these mimics may be integrated with the human immune response by fusing the constant region of an IgG ($F_c$) with a $^{10}$Fn3 module, preferably through the C-terminus of $^{10}$Fn3. The $F_c$ in such a $^{10}$Fn3-$F_c$ fusion molecule activates the complement component of the immune response and increases the therapeutic value of the antibody mimic. Similarly, a fusion

between $^{10}$Fn3 and a complement protein, such as C1q, may be used to target cells, and a fusion between $^{10}$Fn3 and a toxin may be used to specifically destroy cells that carry a particular antigen. In addition, $^{10}$Fn3 in any form may be fused with albumin to increase its half-life in the bloodstream and its tissue penetration. Any of these fusions may be generated by standard techniques, for example, by expression of the fusion protein from a recombinant fusion gene constructed using publically available gene sequences.

Fibronectin Scaffold Multimers

In addition to fibronectin monomers, any of the fibronectin constructs described herein may be generated as dimers or multimers of $^{10}$Fn3-based antibody mimics as a means to increase the valency and thus the avidity of antigen binding. Such multimers may be generated through covalent binding between individual $^{10}$Fn3 modules, for example, by imitating the natural $^{8}$Fn3-$^{9}$Fn3-$^{10}$Fn3 C-to-N-terminus binding or by imitating antibody dimers that are held together through their constant regions. A $^{10}$Fn3-Fc construct may be exploited to design dimers of the general scheme of $^{10}$Fn3-Fc::Fc-$^{10}$Fn3. The bonds engineered into the Fc::Fc interface may be covalent or non-covalent. In addition, dimerizing or multimerizing partners other than Fc can be used in $^{10}$Fn3 hybrids to create such higher order structures.

In particular examples, covalently bonded multimers may be generated by constructing fusion genes that encode the multimer or, alternatively, by engineering codons for cysteine residues into monomer sequences and allowing disulfide bond formation to occur between the expression products. Non-covalently bonded multimers may also be generated by a variety of techniques. These include the introduction, into monomer sequences, of codons corresponding to positively and/or negatively charged

residues and allowing interactions between these residues in the expression products (and therefore between the monomers) to occur. This approach may be simplified by taking advantage of charged residues naturally present in a monomer subunit, for example, the negatively charged residues of fibronectin. Another means for generating non-covalently bonded antibody mimics is to introduce, into the monomer gene (for example, at the amino- or carboxy-termini), the coding sequences for proteins or protein domains known to interact. Such proteins or protein domains include coil-coil motifs, leucine zipper motifs, and any of the numerous protein subunits (or fragments thereof) known to direct formation of dimers or higher order multimers.

Fibronectin-Like Molecules

Although $^{10}$Fn3 represents a preferred scaffold for the generation of antibody mimics, other molecules may be substituted for $^{10}$Fn3 in the molecules described herein. These include, without limitation, human fibronectin modules $^{1}$Fn3-$^{9}$Fn3 and $^{11}$Fn3-$^{17}$Fn3 as well as related Fn3 modules from non-human animals and prokaryotes. In addition, Fn3 modules from other proteins with sequence homology to $^{10}$Fn3, such as tenascins and undulins, may also be used. Modules from different organisms and parent proteins may be most appropriate for different applications; for example, in designing an antibody mimic, it may be most desirable to generate that protein from a fibronectin or fibronectin-like molecule native to the organism for which a therapeutic or diagnostic molecule is intended.

Directed Evolution of Scaffold-Based Binding Proteins

The antibody mimics described herein may be used in any technique for evolving new or improved binding proteins. In one particular example, the target of binding is immobilized on a solid support, such as a column resin or

microtiter plate well, and the target contacted with a library of candidate scaffold-based binding proteins. Such a library may consist of $^{10}$Fn3 clones constructed from the wild type $^{10}$Fn3 scaffold through randomization of the sequence and/or the length of the $^{10}$Fn3 CDR-like loops. If desired, this library may be an RNA-protein fusion library generated, for example, by the techniques described in Szostak et al., U.S.S.N. 09/007,005 and 09/247,190; Szostak et al., WO98/31700; and Roberts & Szostak, Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 12297-12302. Alternatively, it may be a DNA-protein library (for example, as described in Lohse, DNA-Protein Fusions and Uses Thereof, U.S.S.N. 60/110,549, filed December 2, 1998 and _____, filed December 2, 1999). The fusion library is incubated with the immobilized target, the support is washed to remove non-specific binders, and the tightest binders are eluted under very stringent conditions and subjected to PCR to recover the sequence information or to create a new library of binders which may be used to repeat the selection process, with or without further mutagenesis of the sequence. A number of rounds of selection may be performed until binders of sufficient affinity for the antigen are obtained.

In one particular example, the $^{10}$Fn3 scaffold may be used as the selection target. For example, if a protein is required that binds a specific peptide sequence presented in a ten residue loop, a single $^{10}$Fn3 clone is constructed in which one of its loops has been set to the length of ten and to the desired sequence. The new clone is expressed *in vivo* and purified, and then immobilized on a solid support. An RNA-protein fusion library based on an appropriate scaffold is then allowed to interact with the support, which is then washed, and desired molecules eluted and re-selected as described above.

Similarly, the $^{10}$Fn3 scaffold may be used to find natural proteins that interact with the peptide sequence displayed in a $^{10}$Fn3 loop. The $^{10}$Fn3 protein is immobilized as described above, and an RNA-protein fusion library is

screened for binders to the displayed loop. The binders are enriched through multiple rounds of selection and identified by DNA sequencing.

In addition, in the above approaches, although RNA-protein libraries represent exemplary libraries for directed evolution, any type of scaffold-based library may be used in the selection methods of the invention.

Use

The antibody mimics described herein may be evolved to bind any antigen of interest. These proteins have thermodynamic properties superior to those of natural antibodies and can be evolved rapidly in vitro. Accordingly, these antibody mimics may be employed in place of antibodies in all areas in which antibodies are used, including in the research, therapeutic, and diagnostic fields. In addition, because these scaffolds possess solubility and stability properties superior to antibodies, the antibody mimics described herein may also be used under conditions which would destroy or inactivate antibody molecules. Finally, because the scaffolds of the present invention may be evolved to bind virtually any compound, these molecules provide completely novel binding proteins which also find use in the research, diagnostic, and therapeutic areas.

Experimental Results

Exemplary scaffold molecules described above were generated and tested, for example, in selection protocols, as follows.

Library construction

A complex library was constructed from three fragments, each of which contained one randomized area corresponding to a CDR-like loop. The fragments were named BC, DE, and FG, based on the names of the

CDR-H-like loops contained within them; in addition to $^{10}$Fn3 and a randomized sequence, each of the fragments contained stretches encoding an N-terminal $His_6$ domain or a C-terminal FLAG peptide tag. At each junction between two fragments (i.e., between the BC and DE fragments or between the DE and FG fragments), each DNA fragment contained recognition sequences for the EarI Type IIS restriction endonuclease. This restriction enzyme allowed the splicing together of adjacent fragments while removing all foreign, non-$^{10}$Fn3, sequences. It also allows for a recombination-like mixing of the three $^{10}$Fn3 fragments between cycles of mutagenesis and selection.

Each fragment was assembled from two overlapping oligonucleotides, which were first annealed, then extended to form the double-stranded DNA form of the fragment. The oligonucleotides that were used to construct and process the three fragments are listed below; the "Top" and "Bottom" species for each fragment are the oligonucleotides that contained the entire $^{10}$Fn3 encoding sequence. In these oligonucleotides designations, "N" indicates A, T, C, or G; and "S" indicates C or G.

**HfnLbcTop (His):**

5'- GG AAT TCC TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA TTT ACA ATT ACA ATG CAT CAC CAT CAC CAT CAC GTT TCT GAT GTT CCG AGG GAC CTG GAA GTT GTT GCT GCG ACC CCC ACC AGC-3' (SEQ ID NO: 1)

**HfnLbcTop (an alternative N-terminus):**

5'- GG AAT TCC TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA TTT ACA ATT ACA ATG GTT TCT GAT GTT CCG AGG GAC CTG GAA GTT GTT GCT GCG ACC CCC ACC AGC-3' (SEQ ID NO: 2)

**HFnLBCBot-flag8:**

5'-AGC GGA TGC CTT GTC GTC GTC GTC CTT GTA GTC GCT CTT CCC TGT TTC TCC GTA AGT GAT CCT GTA ATA TCT (SNN)7 CCA GCT GAT CAG TAG GCT GGT GGG GGT CGC AGC -3' (SEQ ID NO: 3)

**HFnBC3'-flag8:**

5'-AGC GGA TGC CTT GTC GTC GTC GTC CTT GTA GTC GCT CTT CCC TGT TTC TCC GTA AGT GAT CC-3' (SEQ ID NO: 4)

**HFnLDETop:**

5'- GG AAT TCC TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA TTT ACA ATT ACA ATG CAT CAC CAT CAC CAT CAC CTC TTC ACA GGA GGA AAT AGC CCT GTC C-3' (SEQ ID NO: 5)

**HFnLDEBot-flag8:**

5'-AGC GGA TGC CTT GTC GTC GTC GTC CTT GTA GTC GCT CTT CGT ATA ATC AAC TCC AGG TTT AAG GCC GCT GAT GGT AGC TGT (SNN)4 AGG CAC AGT GAA CTC CTG GAC AGG GCT ATT TCC TCC TGT -3' (SEQ ID NO: 6)

**HFnDE3'-flag8:**

5'-AGC GGA TGC CTT GTC GTC GTC GTC CTT GTA GTC GCT CTT CGT ATA ATC AAC TCC AGG TTT AAG G-3' (SEQ ID NO: 7)

**HFnLFGTop:**

5'- GG AAT TCC TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA TTT ACA ATT ACA ATG CAT CAC CAT CAC CAT CAC CTC TTC TAT ACC ATC ACT GTG TAT GCT GTC-3' (SEQ ID NO: 8)

**HFnLFGBot-flag8:**

5'-AGC GGA TGC CTT GTC GTC GTC GTC CTT GTA GTC TGT TCG GTA ATT AAT GGA AAT TGG (SNN)10 AGT GAC AGC ATA CAC AGT GAT GGT ATA -3' (SEQ ID NO: 9)

**HFnFG3'-flag8:**

5'-AGC GGA TGC CTT GTC GTC GTC GTC CTT GTA GTC TGT TCG GTA ATT AAT GGA AAT TGG -3' (SEQ ID NO: 10)

**T7Tmv (introduces T7 promoter and TMV untranslated region needed for in vitro translation):**

5'- GCG TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA TTT ACA ATT ACA-3' (SEQ ID NO: 11)

**ASAflag8:**

5'-AGC GGA TGC CTT GTC GTC GTC GTC CTT GTA GTC-3' (SEQ ID NO: 12)

**Unispl-s (spint oligonucleotide used to ligate mRNA to the puromycin-containing linker, described by Roberts et al, 1997, supra):**

5'-TTTTTTTTTNAGCGGATGC-3' (SEQ ID NO: 13)

**A18—2PEG (DNA-puromycin linker):**

5'-(A)18(PEG)2CCPur (SEQ ID NO: 14)

The pairs of oligonucleotides (500 pmol of each) were annealed in 100 µL of 10 mM Tris 7.5, 50 mM NaCl for 10 minutes at 85°C, followed by a slow (0.5-1 hour) cooling to room temperature. The annealed fragments with

single-stranded overhangs were then extended using 100 U Klenow (New England Biolabs, Beverly, MA) for each 100 µL aliquot of annealed oligos, and the buffer made of 838.5 µl $H_2O$, 9 µl 1 M Tris 7.5, 5 µl 1M $MgCl_2$, 20 µl 10 mM dNTPs, and 7.5 µl 1M DTT. The extension reactions proceeded for 1 hour at 25°C.

Next, each of the double-stranded fragments was transformed into a RNA-protein fusion (PROfusion™) using the technique developed by Szostak et al., U.S.S.N. 09/007,005 and U.S.S.N. 09/247,190; Szostak et al., WO98/31700; and Roberts & Szostak, Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 12297-12302. Briefly, the fragments were transcribed using an Ambion in vitro transcription kit, MEGAshortscript (Ambion, Austin, TX), and the resulting mRNA was gel-purified and ligated to a DNA-puromycin linker using DNA ligase. The mRNA-DNA-puromycin molecule was then translated using the Ambion rabbit reticulocyte lysate-based translation kit. The resulting mRNA-DNA-puromycin-protein PROfusion™ was purified using Oligo(dT) cellulose, and a complementary DNA strand was synthesized using reverse transcriptase and the RT primers described above (Unisplint-S or flagASA), following the manufacturer's instructions.

The PROfusion™ obtained for each fragment was next purified on the resin appropriate to its peptide purification tag, i.e., on Ni-NTA agarose for the $His_6$-tag and M2 agarose for the FLAG-tag, following the procedure recommended by the manufacturer. The DNA component of the tag-binding PROfusions™ was amplified by PCR using Pharmacia Ready-to-Go PCR Beads, 10 pmol of 5' and 3' PCR primers, and the following PCR program (Pharmacia, Piscataway, NJ): Step 1: 95°C for 3 minutes; Step 2: 95°C for 30 seconds, 58/62°C for 30 seconds, 72°C for 1 minute, 20/25/30 cycles, as required; Step 3: 72°C for 5 minutes; Step 4: 4°C until end.

The resulting DNA was cleaved by 5 U EarI (New England Biolabs)

per1 ug DNA; the reaction took place in T4 DNA Ligase Buffer (New England Biolabs) at 37°C, for 1 hour, and was followed by an incubation at 70°C for15 minutes to inactivate Ear I. Equal amounts of the BC, DE, and FG fragments were combined and ligated to form a full-length $^{10}$Fn3 gene with randomized loops. The ligation required 10 U of fresh EarI (New England Biolabs) and 20 U of T4 DNA Ligase (Promega, Madison, WI), and took 1 hour at 37°C.

Three different libraries were made in the manner described above. Each contained the form of the FG loop with 10 randomized residues. The BC and the DE loops of the first library bore the wild type $^{10}$Fn3 sequence; a BC loop with 7 randomized residues and a wild type DE loop made up the second library; and a BC loop with 7 randomized residues and a DE loop with 4 randomized residues made up the third library. The complexity of the FG loop in each of these three libraries was $10^{13}$; the further two randomized loops provided the potential for a complexity too large to be sampled in a laboratory.

The three libraries constructed were combined into one master library in order to simplify the selection process; target binding itself was expected to select the most suitable library for a particular challenge. PROfusions™ were obtained from the master library following the general procedure described in Szostak et al., U.S.S.N. 09/007,005 and 09/247,190; Szostak et al., WO98/31700; and Roberts & Szostak, Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 12297-12302 (Figure 8).

Fusion Selections

The master library in the PROfusion™ form was subjected to selection for binding to TNF-α. Two protocols were employed: one in which the target was immobilized on an agarose column and one in which the target was immobilized on a BIACORE chip. First, an extensive optimization of conditions to minimize background binders to the agarose column yielded the

favorable buffer conditions of 50 mM HEPES pH 7.4, 0.02% Triton, 100 μg/ml Sheared Salmon Sperm DNA. In this buffer, the non-specific binding of the $^{10}$Fn3 RNA fusion to TNF-α Sepharose was 0.3%. The non-specific binding background of the $^{10}$Fn3 RNA-DNA to TNF-α Sepharose was found to be 0.1%.

During each round of selection on TNF-α Sepharose, the Profusion™ library was first preincubated for an hour with underivatized Sepharose to remove any remaining non-specific binders; the flow-through from this pre-clearing was incubated for another hour with TNF-α Sepharose. The TNF-α Sepharose was washed for 3-30 minutes.

After each selection, the PROfusion™ DNA that had been eluted from the solid support with 0.3 M NaOH or 0.1M KOH was amplified by PCR; a DNA band of the expected size persisted through multiple rounds of selection (Figure 9); similar results were observed in the two alternative selection protocols, and only the data from the agarose column selection is shown in Figure 9.

In the first seven rounds, the binding of library PROfusions™ to the target remained low; in contrast, when free protein was translated from DNA pools at different stages of the selection, the proportion of the column binding species increased significantly between rounds (Figure 10). Similar selections may be carried out with any other binding species target (for example, IL-1 and IL-13).

Animal Studies

Wild-type $^{10}$Fn3 contains an integrin-binding tripepetide motif, Arginine 78 - Glycine 79 - Aspartate 80 (the "RGD motif') at the tip of the FG loop. In order to avoid integrin binding and a potential inflammatory response based on this tripeptide in vivo, a mutant form of $^{10}$Fn3 was generated that

contained an inert sequence, Serine 78 - Glycine 79 - Glutamate 80 (the "SGE mutant"), a sequence which is found in the closely related, wild-type $^{11}$Fn3 domain. This SGE mutant was expressed as an N-terminally $His_6$-tagged, free protein in E. coli, and purified to homogeneity on a metal chelate column followed by a size exclusion column.

In particular, the DNA sequence encoding $His_6$-$^{10}$Fn3(SGE) was cloned into the pET9a expression vector and transformed into BL21 DE3 pLysS cells. The culture was then grown in LB broth containing 50 µg/mL kanamycin at 37°C, with shaking, to $A_{560}$=1.0, and was then induced with 0.4 mM IPTG. The induced culture was further incubated, under the same conditions, overnight (14-18 hours); the bacteria were recovered by standard, low speed centrifugation. The cell pellet was resuspended in 1/50 of the original culture volume of lysis buffer (50 mM Tris 8.0, 0.5 M NaCl, 5% glycerol, 0.05% Triton X-100, and 1 mM PMSF), and the cells were lysed by passing the resulting paste through a Microfluidics Corporation Microfluidizer M110-EH, three times. The lysate was clarified by centrifugation, and the supernatant was filtered through a 0.45 µm filter followed by filtration through a 0.2 µm filter. 100 mL of the clarified lysate was loaded onto a 5 mL Talon cobalt column (Clontech, Palo Alto, CA), washed by 70 mL of lysis buffer, and eluted with a linear gradient of 0-30 mM imidazole in lysis buffer. The flow rate through the column through all the steps was 1 mL/min. The eluted protein was concentrated 10-fold by dialysis (MW cutoff = 3,500) against 15,000-20,000 PEG. The resulting sample was dialysed into buffer 1 (lysis buffer without the glycerol), then loaded, 5 mL at a time, onto a 16 x 60 mm Sephacryl 100 size exclusion column equilibrated in buffer 1. The column was run at 0.8 mL/min, in buffer 1; all fractions that contained a protein of the expected MW were pooled, concentrated 10X as described above, then dialyzed into PBS. Toxikon (MA) was engaged to perform endotoxin screens

and animal studies on the resulting sample.

In these animal studies, the endotoxin levels in the samples examined to date have been below the detection level of the assay. In a preliminary toxicology study, this protein was injected into two mice at the estimated 100X therapeutic dose of 2.6 mg/mouse. The animals survived the two weeks of the study with no apparent ill effects. These results suggest that $^{10}$Fn3 may be incorporated safely into an IV drug.

Alternative Constructs for In Vivo Use

To extend the half life of the 8 kD $^{10}$Fn3 domain, a larger molecule has also been constructed that mimics natural antibodies. This $^{10}$Fn3-$F_c$ molecule contains the -$CH_1$-$CH_2$-$CH_3$ (Figure 11) or -$CH_2$-$CH_3$ domains of the IgG constant region of the host; in these constructs, the $^{10}$Fn3 domain is grafted onto the N-terminus in place of the IgG $V_H$ domain (Figures 11 and 12). Such antibody-like constructs are expected to improve the pharmacokinetics of the protein as well as its ability to harness the natural immune response.

In order to construct the murine form of the $^{10}$Fn3-$CH_1$-$CH_2$-$CH_3$ clone, the -$CH_1$-$CH_2$-$CH_3$ region was first amplified from a mouse liver spleen cDNA library (Clontech), then ligated into the pET25b vector. The primers used in the cloning were 5' Fc Nest and 3' 5 Fc Nest, and the primers used to graft the appropriate restriction sites onto the ends of the recovered insert were 5' Fc HIII and 3' Fc Nhe:

5' Fc Nest 5'GCG GCA GGG TTT GCT TAC TGG GGC CAA GGG 3' (SEQ ID NO: 15);

3' Fc Nest 5'GGG AGG GGT GGA GGT AGG TCA CAG TCC 3' (SEQ ID NO: 16);

3' Fc Nhe 5' TTT GCT AGC TTT ACC AGG AGA GTG GGA GGC 3' (SEQ

ID NO: 17); and

5' Fc HIII 5' AAA AAG CTT GCC AAA ACG ACA CCC CCA TCT GTC 3' (SEQ ID NO: 18).

Further PCR is used to remove the $CH_1$ region from this clone and create the Fc part of the shorter, $^{10}Fn3$-$CH_2$-$CH_3$ clone. The sequence encoding $^{10}Fn3$ is spliced onto the 5' end of each clone; either the wild type $^{10}Fn3$ cloned from the same mouse spleen cDNA library or a modified $^{10}Fn3$ obtained by mutagenesis or randomization of the molecules can be used. The oligonucleotides used in the cloning of murine wild-type $^{10}Fn3$ were:

Mo 5PCR-NdeI:

5' CATATGGTTTCTGATATTCCGAGAGATCTGGAG 3' (SEQ ID NO: 19);

Mo5PCR-His-NdeI (for an alternative N-terminus with the $His_6$ purification tag):

5' CAT ATG CAT CAC CAT CAC CAT CAC GTT TCT GAT ATT CCG AGA G 3' (SEQ ID NO: 20); and

Mo3PCR-EcoRI: 5' GAATTCCTATGTTTTATAATTGATGGAAAC3' (SEQ ID NO: 21).

The human equivalents of the clones are constructed using the same strategy with human oligonucleotide sequences.

Other embodiments are within the claims.

All publications, patents, and patent applications mentioned herein are hereby incorporated by reference.

What is claimed is:

## Claims

1. A protein comprising a fibronectin type III domain having at least one randomized loop, said protein being characterized by its ability to bind to a compound that is not bound by the corresponding naturally-occurring fibronectin.

2. The protein of claim 1, wherein said fibronectin type III domain is a mammalian fibronectin type III domain.

3. The protein of claim 2, wherein said fibronectin type III domain is a human fibronectin type III domain.

4. The protein of claim 1, wherein said protein comprises the tenth module of said fibronectin type III domain ($^{10}$Fn3).

5. The protein of claim 4, wherein said compound binding is mediated by one $^{10}$Fn3 loop.

6. The protein of claim 4, wherein said compound binding is mediated by two $^{10}$Fn3 loops.

7. The protein of claim 4, wherein said compound binding is mediated by three $^{10}$Fn3 loops.

8. The protein of claim 4, wherein the second loop of said $^{10}$Fn3 is extended in length relative to the naturally-occurring module.

9. The protein of claim 4, wherein said $^{10}$Fn3 lacks an integrin-binding motif.

10. The protein of claim 9, wherein said integrin-binding motif is replaced by an amino acid sequence comprising a basic amino acid-neutral amino acid-acidic amino acid motif.

11. The protein of claim 10, wherein said integrin-binding motif is replaced by an amino acid sequence comprising serine-glycine-glutamate.

12. The protein of claim 1, wherein said protein lacks disulfide bonds.

13. The protein of claim 1, wherein said protein is part of a fusion protein.

14. The protein of claim 13, wherein said fusion protein further comprises an immunoglobulin $F_c$ domain.

15. The protein of claim 13, wherein said fusion protein further comprises a complement protein.

16. The protein of claim 13, wherein said fusion protein further comprises a toxin protein.

17. The protein of claim 13, wherein said fusion protein further comprises an albumin protein.

18. The protein of claim 1, wherein said protein is covalently bound to a nucleic acid.

19. The protein of claim 18, wherein said nucleic acid encodes said protein.

20. The protein of claim 18, wherein said nucleic acid is RNA.

21. The protein of claim 1, wherein said protein is a multimer.

22. The protein of claim 1 or 9, wherein said protein is formulated in a physiologically-acceptable carrier.

23. A nucleic acid encoding the protein of claim 1 or 4.

24. The nucleic acid of claim 23, wherein said nucleic acid is DNA.

25. The nucleic acid of claim 23, wherein said nucleic acid is RNA.

26. A method for generating a protein comprising a fibronectin type III domain which is pharmaceutically acceptable to a mammal, said method comprising removing an integrin-binding domain from said fibronectin type III domain.

27. The method of claim 26, wherein said integrin binding motif is replaced by an amino acid sequence comprising a basic amino acid-neutral amino acid-acidic amino acid motif.

28. The protein of claim 27, wherein said integrin-binding motif is replaced by an amino acid sequence comprising serine-glycine-glutamate.

29. The method of claim 26, wherein said at least one loop of said fibronectin type III domain is randomized.

30. The method of claim 26, wherein said protein comprises the tenth module of said fibronectin type III domain.

31. The protein of claim 26, wherein said protein is part of a fusion protein.

32. The protein of claim 31, wherein said fusion protein further comprises an immunoglobulin $F_c$ domain.

33. The protein of claim 31, wherein said fusion protein further comprises a complement protein.

34. The protein of claim 31, wherein said fusion protein further comprises a toxin protein.

35. The protein of claim 31, wherein said fusion protein further comprises an albumin protein.

36. The method of claim 26, wherein said mammal is a human.

37. A method for obtaining a protein which binds to a compound, said method comprising:

(a) contacting said compound with a candidate protein, said candidate protein comprising a fibronectin type III domain having at least one randomized loop, said contacting being carried out under conditions that allow compound-protein complex formation; and

(b) obtaining, from said complex, said protein which binds to said compound.

38. A method for obtaining a compound which binds to a protein, said protein comprising a fibronectin type III domain having at least one randomized loop, said method comprising:

(a) contacting said protein with a candidate compound, said contacting being carried out under conditions that allow compound-protein complex formation; and

(b) obtaining, from said complex, said compound which binds to said protein.

39. The method of claim 37, said method further comprising randomizing at least one loop of said fibronectin type III domain of said protein obtained in step (b) and repeating said steps (a) and (b) using said further randomized protein.

40. The method of claim 38, said method further comprising modifying said compound obtained in step (b) and repeating said steps (a) and (b) using said further modified compound.

41. The method of claim 37 or 38, wherein said compound is a protein.

42. The method of claim 37 or 38, wherein said fibronectin type III domain is a mammalian fibronectin type III domain.

43. The method of claim 42, wherein said fibronectin type III domain is a human fibronectin type III domain.

44. The method of claim 37 or 38, wherein said protein comprises the tenth module of said fibronectin type III domain ($^{10}$Fn3).

45. The method of claim 44, wherein said compound binding is mediated by one $^{10}$Fn3 loop.

46. The method of claim 44, wherein said compound binding is mediated by two $^{10}$Fn3 loops.

47. The method of claim 44, wherein said compound binding is mediated by three $^{10}$Fn3 loops.

48. The method of claim 44, wherein the second loop of said $^{10}$Fn3 is extended in length relative to the naturally-occurring module.

49. The method of claim 44, wherein said $^{10}$Fn3 lacks an integrin-binding motif.

50. The method of claim 37, wherein said compound is immobilized on a solid support.

51. The method of claim 38, wherein said protein is immobilized on

a solid support.

52. The method of claim 50 or 51, wherein said solid support is a column or microchip.

FIG. 1

FIG. 2

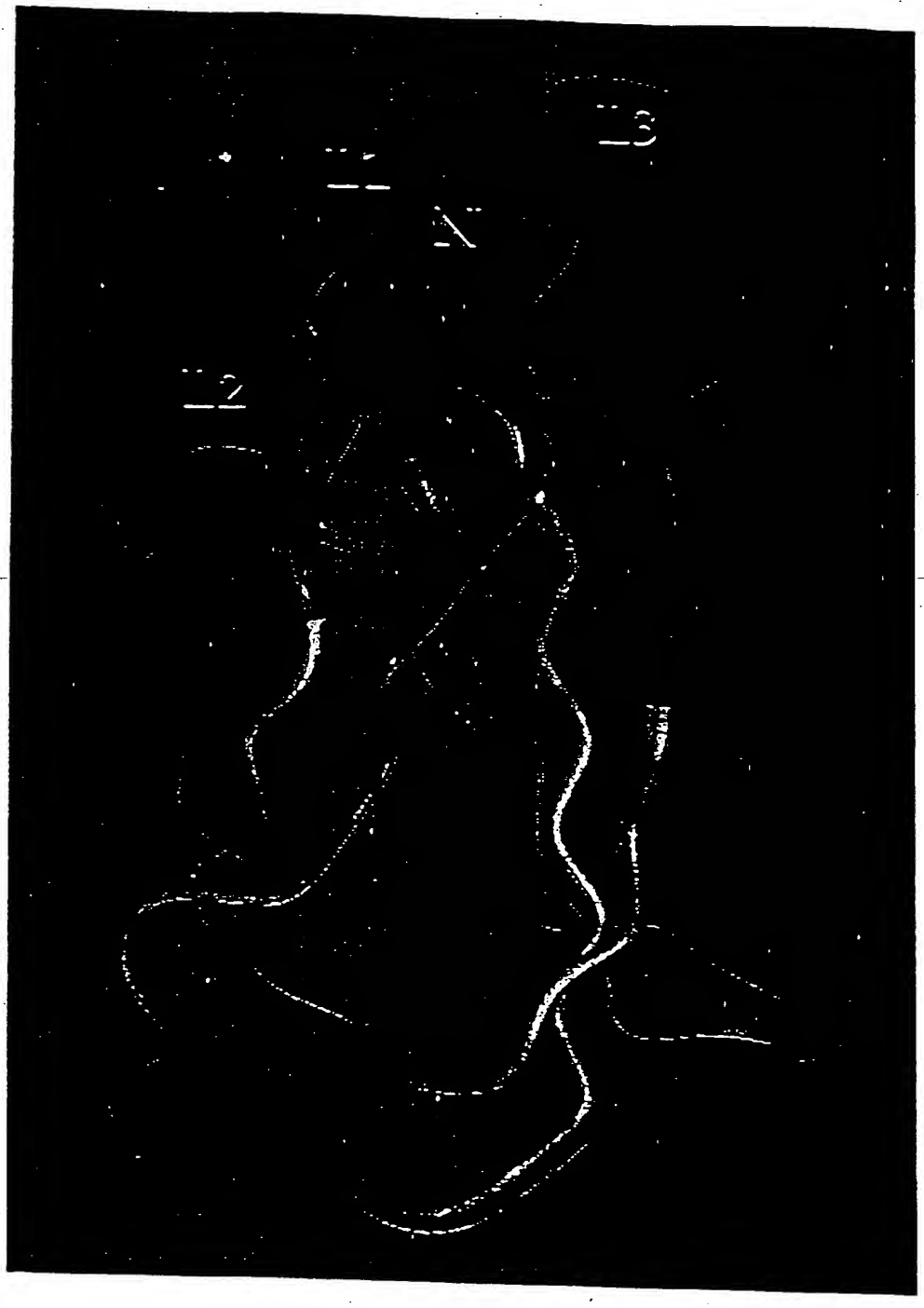
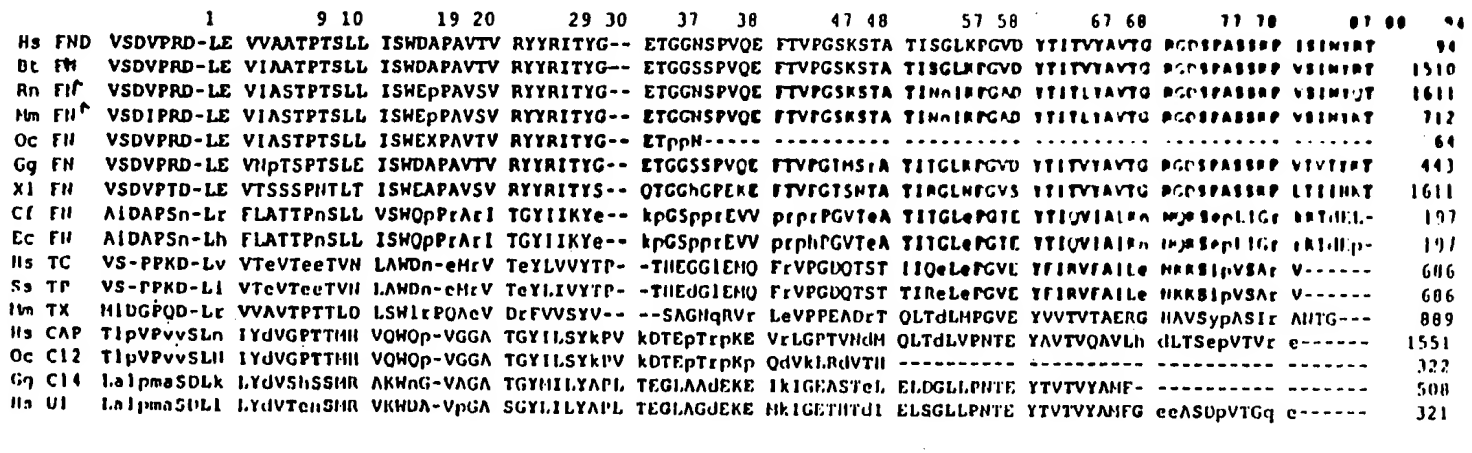



FIG. 3



```
                  1        9 10      19 20      29 30    37  38      47 48      57 58      67 68      77 78      87 88   94
Hs FND  VSDVPRD-LE VVAATPTSLL ISWDAPAVTV RYYRITYG-- ETGGNSPVQE FTVPGSKSTA TISGLKPGVD YTITVYAVTG RGDSPASSKP ISINYRT    94
Bt FN   VSDVPRD-LE VIAATPTSLL ISWDAPAVTV RYYRITYG-- ETGGSSPVQE FTVPGSKSTA TISGLKPGVD YTITVYAVTG RGDSPASSKP VSINYRT  1510
Rn FN   VSDVPRD-LE VIASTPTSLL ISWEPPAVSV RYYRITYG-- ETGGNSPVQE FTVPGSKSTA TINGLRPGAD YTITLYAVTG RGDSPASSKP VSINYQT  1611
Mm FN   VSDIPRD-LE VIASTPTSLL ISWEPPAVSV RYYRITYG-- ETGGNSPVQE FTVPGSKSTA TINGLRPGAD YTITLYAVTG RGDSPASSKP VSINYRT   712
Oc FN   VSDVPRD-LE VIASTPTSLL ISWEXPAVTV RYYRITYG-- ETppN----- ---------- ---------- ---------- ---------- -------    64
Gg FN   VSDVPRD-LE VNpTSPTSLE ISWDAPAVTV RYYRITYG-- ETGGSSPVQE FTVPGTMSTA TINGLKPGVD YTITVYAVTG RGDSPASSKP VTVTYRT   443
Xl FN   VSDVPTD-LE VTSSSPNTLT ISWEAPAVSV RYYRITYS-- QTGGhGPEKE FTVPGTSNTA TINGLNPGVS YTITVYAVTG RGDSPASSKP LTFINRT  1611
Cf FN   AIDAPSn-Lr FLATTPnSLL VSWQpPrArI TGYIIKYe-- kpGSpprEVV prprPGVTeA TITGLePGIE YTIQVIAlRn wpSepLIGr kRTdELp-   197
Ec FN   AIDAPSn-Lh FLATTPnSLL ISWQpPrArI TGYIIKYe-- kpGSpprEVV prphPGVTeA TITGLePGIE YTIQVIAlRn wpSepLIGr kRTdELp-   197
Hs TC   VS-PPKD-Lv VTeVTeeTVN LAWDn-eHrV TeYLVVYTP- -THEGGlEHQ FrVPGDQTST TIQeLePGVE YFIRVFAILe HRKSlpVSAr V------   606
Ss TP   VS-PPKD-Ll VTeVTeeTVN LAWDn-eHrV TeYLIVYTP- -THEdGlEHQ FrVPGDQTST TIReLePGVE YFIRVFAILe HKKSlpVSAr V------   606
Hm TX   HIDGPQD-Lr VVAVTPTTLD LSWIrPQAeV DrFVVSYV-- --SAGHqRVr LeVPPEADrT QLTdLHPGVE YVVTVTAERG HAVSypASIr ANTG---   889
Hs CAP  TIpVPvvSLn IYdVGPTTHH VQWQp-VGGA TGYILSYkPV kDTEpTrpKE VrLGPTVNdM QLTdLVPNTE YAVTVQAVLh dLTSepVTVr e------  1551
Oc C12  TIpVPvvSLH IYdVGPTTHH VQWQp-VGGA TGYILSYkPV kDTEpTrpKp QdVkLRdVTH ---------- ---------- ---------- -------   322
Gg C14  LalpmaSDLk LYdVShSSHR AKWnG-VAGA TGYHILYAPL TEGLAAdEKE IkIGEASTeL ELDGLLPNTE YTVTVYAMF- ---------- -------   508
Hs UI   LalpmaSDLl LYdVTenSHR VKHDA-VpGA SGYLILYAPL TEGLAGdEKE NkIGETHTdI ELSGLLPNTE YTVTVYAMFG eeASDpVTGq e------   321

var.    --- -- -  -  --      -----  -    -  ---------  - - - -    - -             - --  --  - -  - -

cons.      P  L  V     SL    W     V   Y I  Y                      I  L PGVD Y ITV A  G    S     P
           H     I     TV          A   F V                        L  I NTE   VQL   N           R
                 L     M           I       L                           AS    R    E
                 F
```

| | | | |
|---|---|---|---|
| CAP | Collagen alpha precursor | BOLD | identical to Hs FND |
| C12 | Collagen type 12 | | |
| FND | Fibronectin type III domain | lower case | non-conservative substitution |
| FN | Fibronectin | | (charge reversal, change between hydrophobic |
| TP | Tenascin precursor | | and charged, addition or removal of P) |
| TC | Tenascin-C | - | position of non-conservative substitutions |
| UI | Undulin 1 | | |

| | | |
|---|---|---|
| Bt | *Bovis taurus* | cow |
| Cf | *Canis familiaris* | dog |
| Ec | *Equus caballis* | horse |
| Ss | *Sus scrofa* | pig |
| Hs | *Homo sapiens* | human |
| Oc | *Oryctolagus cuniculus* | rabbit |
| Xl | *Xenupus laevis* | African clawed frog |

FIG. 4

FIG. 5

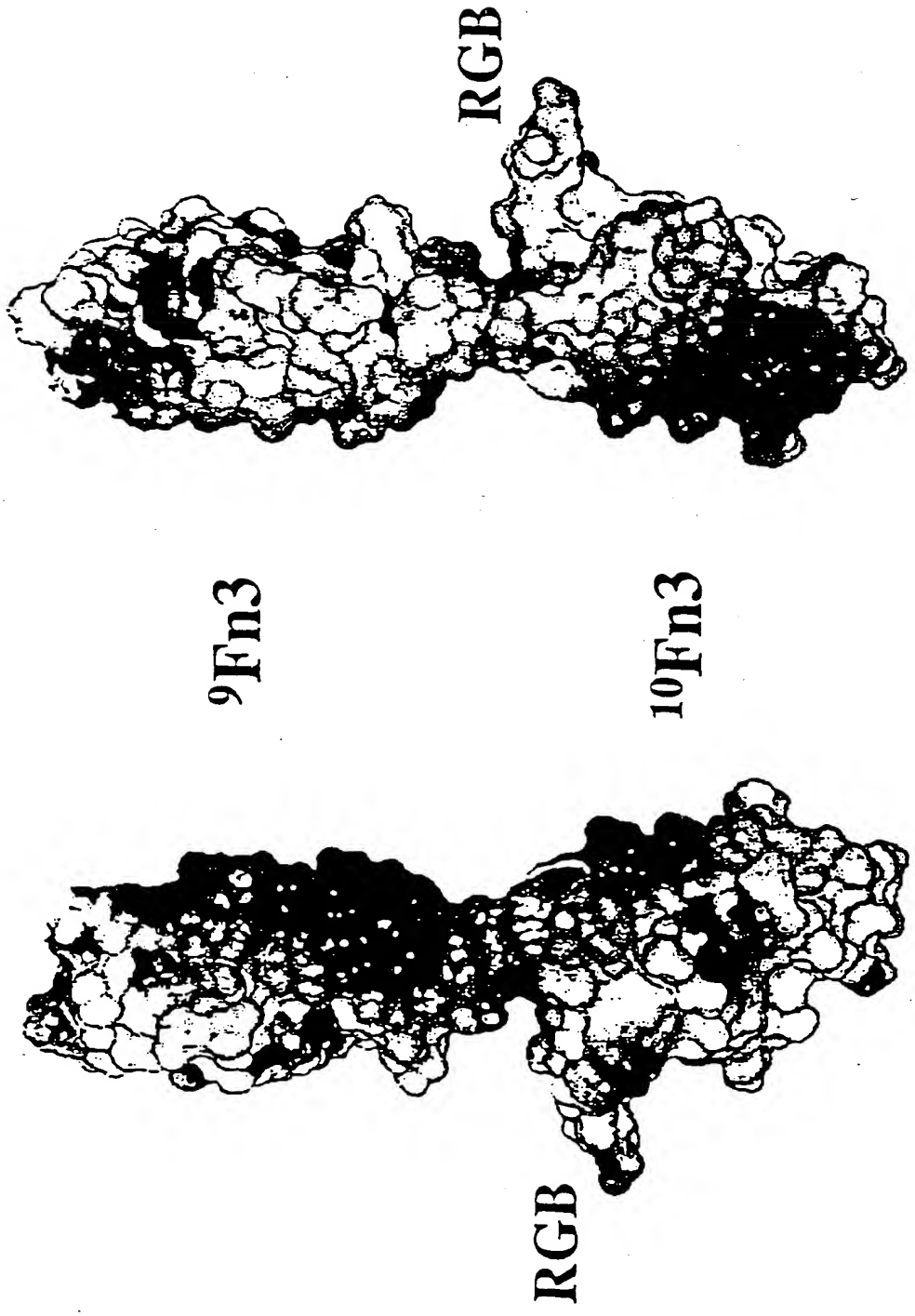



FIG. 6

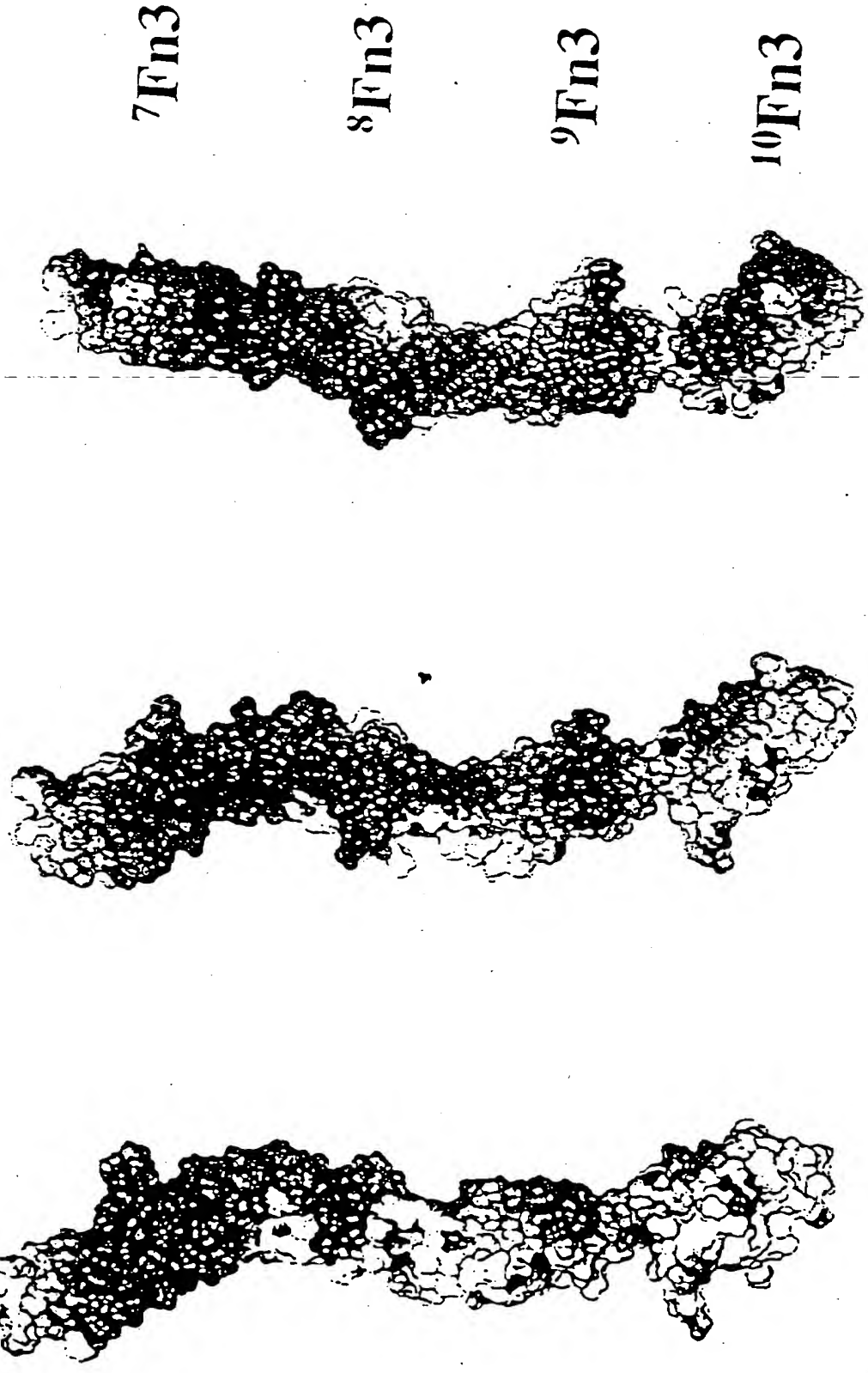



FIG. 7

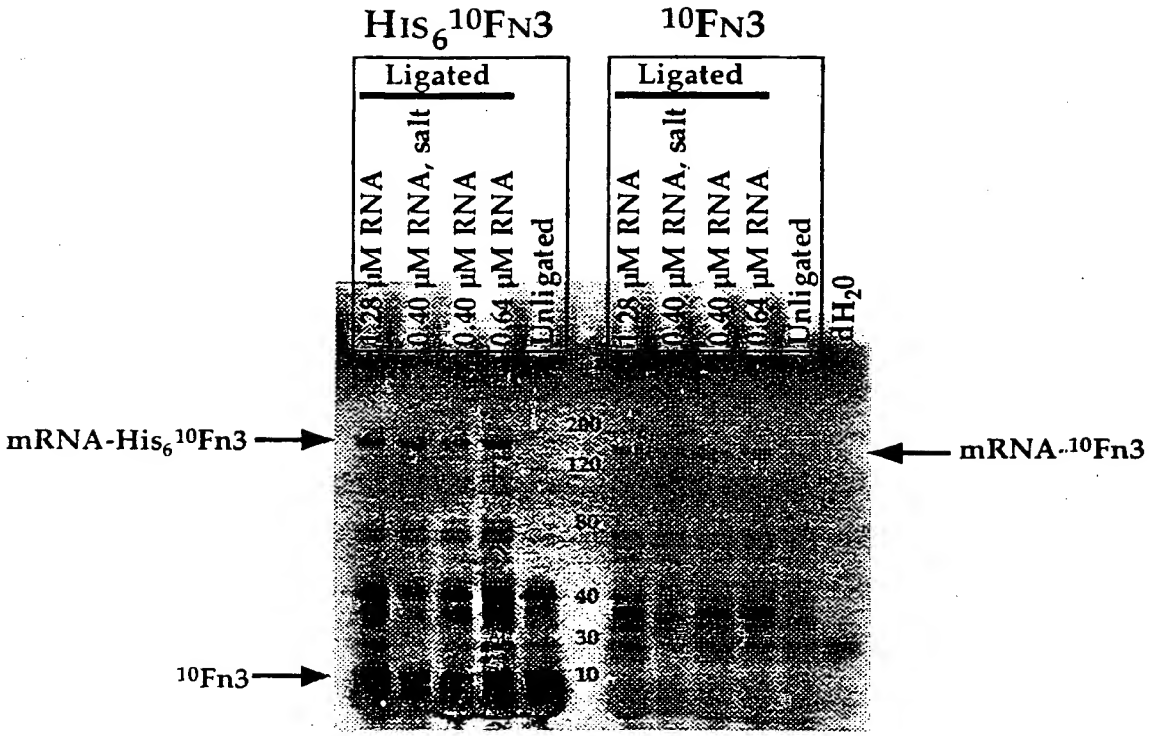

His6 10Fn3
10Fn3
Ligated
Ligated
1.28 µM RNA
0.40 µM RNA, salt
0.40 µM RNA
0.64 µM RNA
Unligated
1.28 µM RNA
0.40 µM RNA, salt
0.40 µM RNA
0.64 µM RNA
Unligated
dH2O
mRNA-His6 10Fn3
mRNA-10Fn3
10Fn3
200
120
80
40
30
10

Figure 8

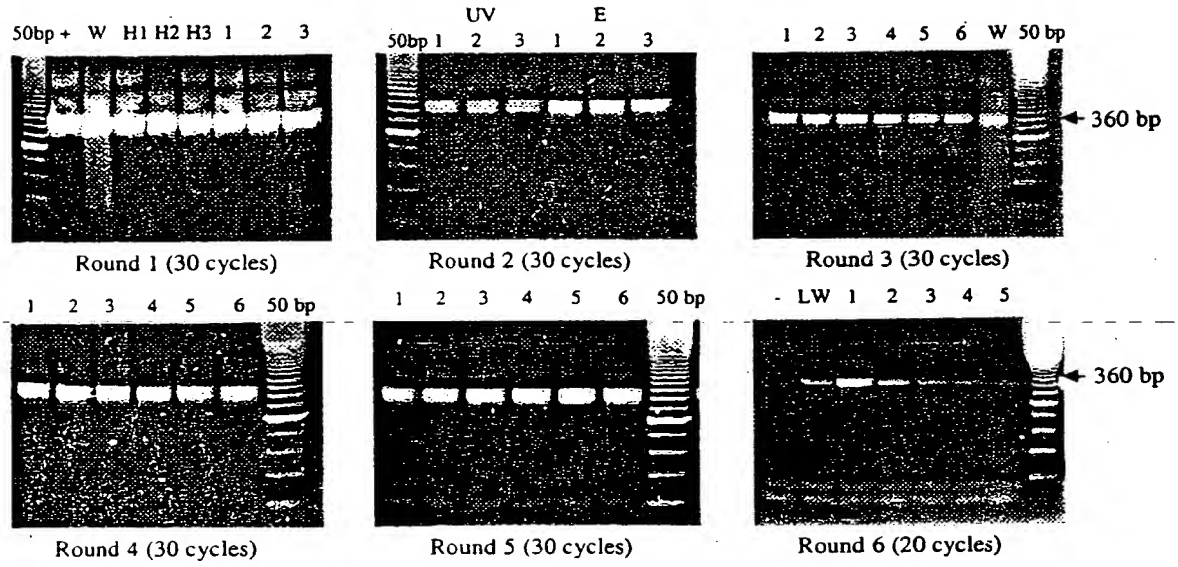

50bp + W H1 H2 H3 1 2 3
Round 1 (30 cycles)
UV E
50bp 1 2 3 1 2 3
Round 2 (30 cycles)
1 2 3 4 5 6 W 50 bp
360 bp
Round 3 (30 cycles)
1 2 3 4 5 6 50 bp
Round 4 (30 cycles)
1 2 3 4 5 6 50 bp
Round 5 (30 cycles)
- LW 1 2 3 4 5
360 bp
Round 6 (20 cycles)

Figure 9

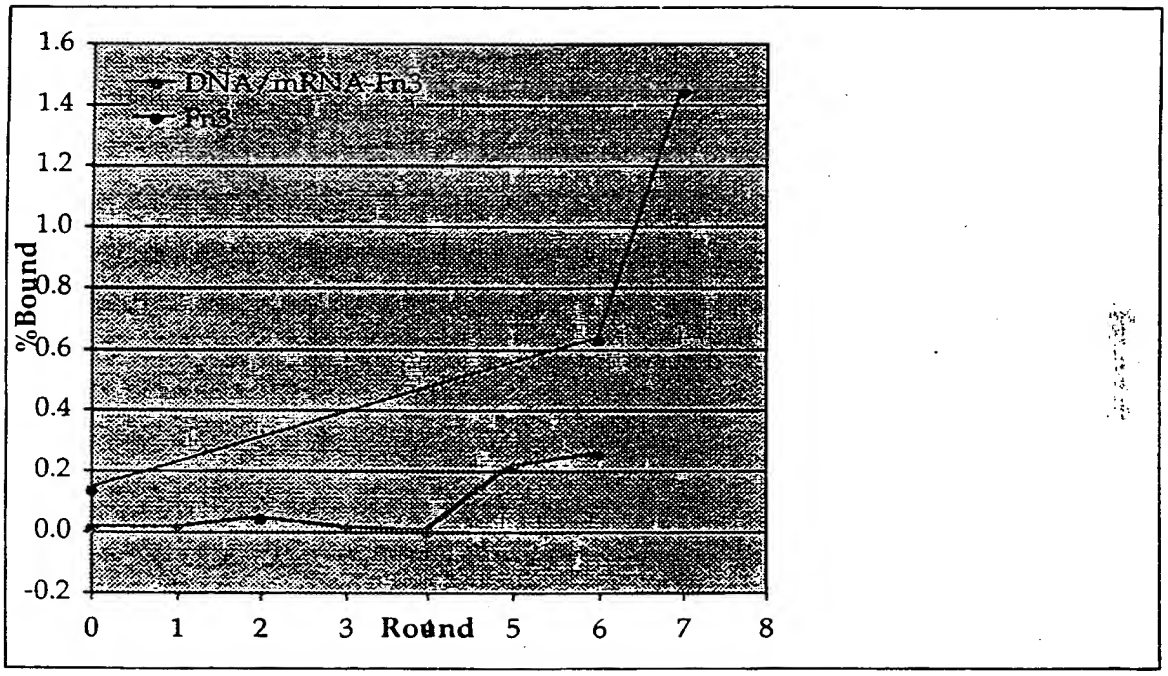
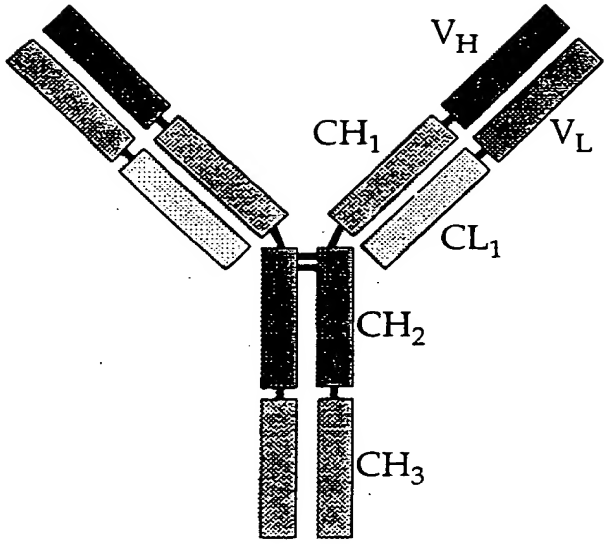
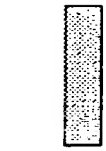

DNA/mRNA-Fn3
Fn3
%Bound
Round
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
0
1
2
3
5
6
7
8

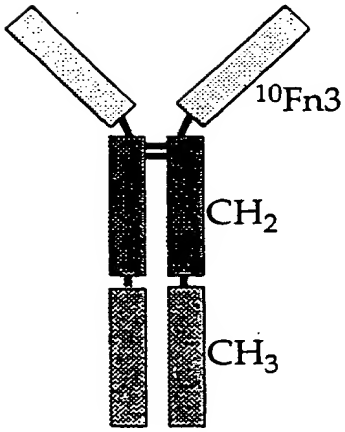
Figure 10



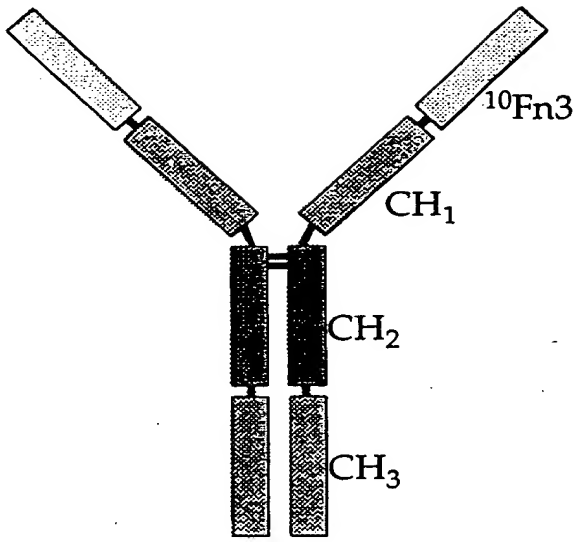

IgG



$^{10}$Fn3




$^{10}$Fn3-$CH_2$-$CH_3$




$^{10}$Fn3-$CH_1$-$CH_2$-$CH_3$

Figure 11

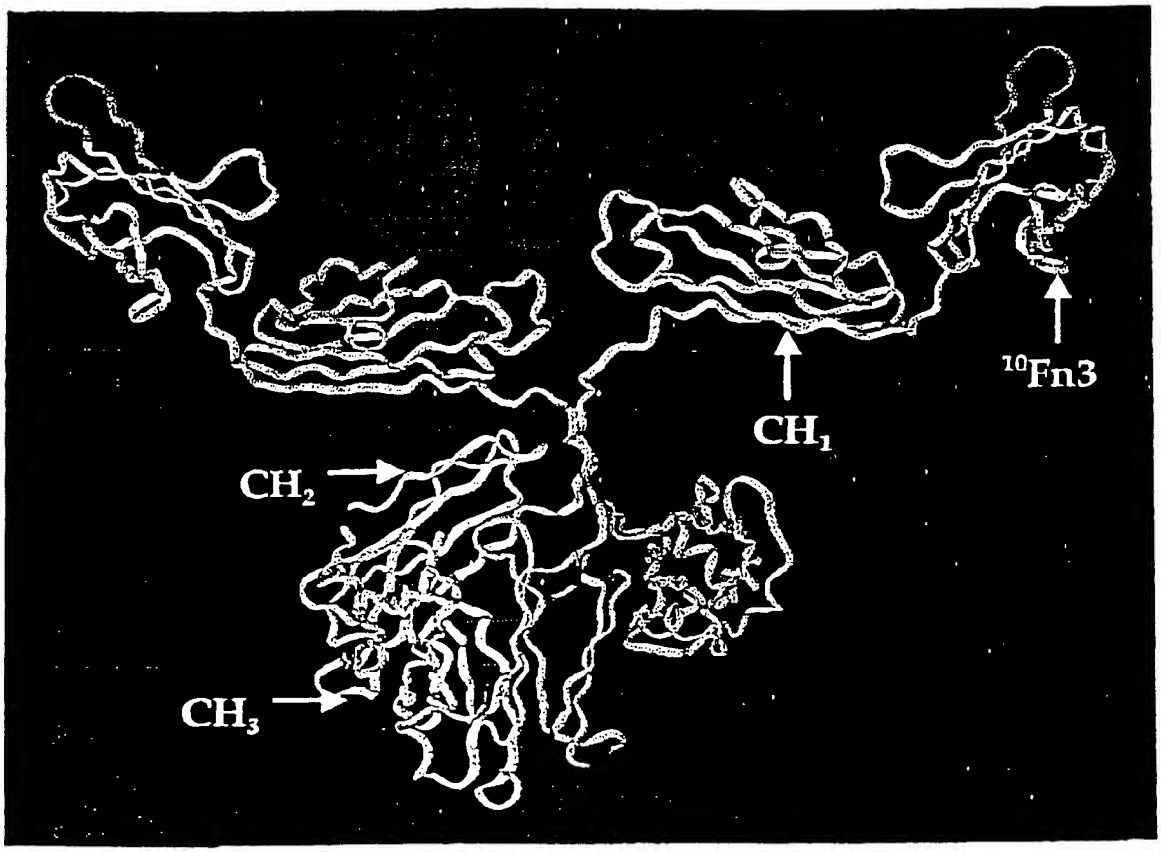



Figure 12

SEQUENCE LISTING

<110> Phylos, Inc.

<120> PROTEIN SCAFFOLDS FOR ANTIBODY MIMICS
AND OTHER BINDING PROTEINS

<130> 50036/021WO2

<150> 60/111,737
<151> 1998-12-10

<160> 21

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 122
<212> DNA
<213> Homo sapiens

<400> 1
ggaattccta atacgactca ctatagggac aattactatt tacaattaca atgcatcacc 60
atcaccatca cgtttctgat gttccgaggg acctggaagt tgttgctgcg acccccacca 120
gc 122

<210> 2
<211> 104
<212> DNA
<213> Homo sapiens

<400> 2
ggaattccta atacgactca ctatagggac aattactatt tacaattaca atggtttctg 60
atgttccgag ggacctggaa gttgttgctg cgacccccac cagc 104

<210> 3
<211> 126
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (1)...(126)
<223> n = A,T,C or G

<221> misc_feature
<222> (1)...(126)
<223> s = C or G

<400> 3
agcggatgcc ttgtcgtcgt cgtccttgta gtcgctcttc cctgtttctc cgtaagtgat 60
cctgtaatat ctsnnsnnsn nsnnsnnsnn snnccaqctg atcagtaggc tggtgggggt 120
cgcagc 126

<210> 4
<211> 62
<212> DNA
<213> Homo sapiens

<400> 4
agcggatgcc ttgtcgtcgt cgtccttgta gtcgctcttc cctgtttctc cgtaagtgat 60
cc 62

<210> 5
<211> 99
<212> DNA
<213> Homo sapiens

<400> 5
ggaattccta tacgactca ctatagggac aattactatt tacaattaca atgcatcacc 60
atcaccatca cctcttcaca ggaggaaata gccctgtcc 99

<210> 6
<211> 132
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (1)...(132)
<223> n = A,T,C or G

<221> misc_feature
<222> (1)...(132)
<223> s = C or G

<400> 6
agcggatgcc ttgtcgtcgt cgtccttgta gtcgctcttc gtataatcaa ctccaggttt 60
aaggccgctg atggtagctg tsnnsnnsnn snnaggcaca gtgaactcct ggacagggct 120
atttcctcct gt 132

<210> 7
<211> 64
<212> DNA
<213> Homo sapiens

<400> 7
agcggatgcc ttgtcgtcgt cgtccttgta gtcgctcttc gtataatcaa ctccaggttt 60
aagg 64

<210> 8
<211> 101
<212> DNA
<213> Homo sapiens

<400> 8
ggaattccta tacgactca ctatagggac aattactatt tacaattaca atgcatcacc 60
atcaccatca cctcttctat accatcactg tgtatgctgt c 101

<210> 9
<211> 114
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (1)...(114)
<223> n = A,T,C or G

<221> misc_feature
<222> (1)...(114)
<223> s = C or G

<400> 9
agcggatgcc ttgtcgtcgt cgtccttgta gtctgttcgg taattaatgg aaattggsnn 60
snnsnnsnns nnsnnsnnsn nsnnsnnagt gacagcatac acagtgatgg tata 114

<210> 10
<211> 57
<212> DNA
<213> Homo sapiens

<400> 10
agcggatgcc ttgtcgtcgt cgtccttgta gtctgttcgg taattaatgg aaattgg 57

<210> 11
<211> 45
<212> DNA
<213> T7 phage and tobacco mosaic virus

<400> 11
gcgtaatacg actcactata gggacaatta ctatttacaa ttaca 45

<210> 12
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> Flag sequence

<400> 12
agcggatgcc ttgtcgtcgt cgtccttgta gtc 33

<210> 13
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Splint oligonucleotide

<221> misc_feature
<222> (1)...(19)

<223> n = A,T,C or G

<400> 13
ttttttttn agcggatgc 19

<210> 14
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Puromycin linker oligonucleotide

<400> 14
aaaaaaaaaa aaaaaaaacc 20

<210> 15
<211> 30
<212> DNA
<213> Mus musculus

<400> 15
gcggtagttt ttacttactg gggccaaggg 30

<210> 16
<211> 27
<212> DNA
<213> Mus musculus

<400> 16
gggaggagtg gaggtaggtc acagtcc 27

<210> 17
<211> 30
<212> DNA
<213> Mus musculus

<400> 17
tttgctagct ttaccaggag agtgggaggc 30

<210> 18
<211> 33
<212> DNA
<213> Mus musculus

<400> 18
aaaaagcttg ccaaaacgac acccccatct gtc 33

<210> 19
<211> 33
<212> DNA
<213> Mus musculus

<400> 19
catatggttt ctgatattcc gagagatctg gag 33

<210> 20
<211> 43
<212> DNA
<213> Mus musculus

<400> 20
catatgcatc accatcacca tcacgtttct gatattccga gag 43

<210> 21
<211> 30
<212> DNA
<213> Mus musculus

<400> 21
gaattcctat gttttataat tgatggaaac 30

# INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/29317

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :G01N 33/536; C07K 14/00; C12N 15/00

US CL :435/7.1, 7.6, 69.1; 436/86, 87, 536; 530/350

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/7.1, 7.6, 69.1; 436/86, 87, 536; 530/350

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| A | MARKLAND et al. Iterative Optimization of High-Affinity Protease Inhibitors Using Phage Display. 1. Plasmin. Biochemistry. 1996, Vol. 35, No. 24, pages 8045-8057. | 1-52 |
| A | NORD et al. Binding Proteins Selected from Combinatorial Libraries of an α-helical bacterial receptor domain. Nature Biotechnology. August 1997, Vol. 15, pages 772-777. | 1-52 |
| A | KU et al. Alternate Protein Framework for Molecular Recognition. Proc. Natl. Acad. Sci. USA. July 1995, Vol. 92, pages 6552-6556. | 1-52 |

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 13 MARCH 2000 | 06 APR 2000 |
| Name and mailing address of the ISA/US<br>Commissioner of Patents and Trademarks<br>Box PCT<br>Washington, D.C. 20231<br>Facsimile No. (703) 305-3230 | Authorized officer<br>HOLLY SCHNIZER<br>Telephone No. (703) 308-0196 |

Form PCT/ISA/210 (second sheet) (July 1998)☆

# INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/29317

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

STN (medline, biosis, embase, caplus), EAST (uspat, derwent, EPO), search terms: fibronectin, fn3, type III domain, randomized, antibody mimics, scaffolds, directed evolution, complementary-determining region, integrin binding, minibody, tenlamistat.

CORRECTED VERSION*

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 :
G01N 33/536, C07K 14/00, C12N 15/00

A1

(11) International Publication Number: WO 00/34784

(43) International Publication Date: 15 June 2000 (15.06.00)

(21) International Application Number: PCT/US99/29317

(22) International Filing Date: 9 December 1999 (09.12.99)

(30) Priority Data:
60/111,737 10 December 1998 (10.12.98) US

(71) Applicant: PHYLOS, INC. [US/US]; 128 Spring Street, Lexington, MA 02421 (US).

(72) Inventor: LIPOVSEK, Dasa; 45 Sunset Road, Cambridge, MA 02138 (US).

(74) Agent: ELBING, Karen; Clark & Elbing LLP, 176 Federal Street, Boston, MA 02110–2214 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Published**
*With international search report.*
*Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.*

(54) Title: PROTEIN SCAFFOLDS FOR ANTIBODY MIMICS AND OTHER BINDING PROTEINS

(57) Abstract

Disclosed herein are proteins that include a fibronectin type III domain having at least one randomized loop. Also disclosed herein are nucleic acids encoding such proteins and the use of such proteins in methods for evolving novel compound–binding species and their ligands.

*(Referred to in PCT Gazette No. 46/2000, Section II)

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **AL** | Albania | **ES** | Spain | **LS** | Lesotho | **SI** | Slovenia |
| **AM** | Armenia | **FI** | Finland | **LT** | Lithuania | **SK** | Slovakia |
| **AT** | Austria | **FR** | France | **LU** | Luxembourg | **SN** | Senegal |
| **AU** | Australia | **GA** | Gabon | **LV** | Latvia | **SZ** | Swaziland |
| **AZ** | Azerbaijan | **GB** | United Kingdom | **MC** | Monaco | **TD** | Chad |
| **BA** | Bosnia and Herzegovina | **GE** | Georgia | **MD** | Republic of Moldova | **TG** | Togo |
| **BB** | Barbados | **GH** | Ghana | **MG** | Madagascar | **TJ** | Tajikistan |
| **BE** | Belgium | **GN** | Guinea | **MK** | The former Yugoslav Republic of Macedonia | **TM** | Turkmenistan |
| **BF** | Burkina Faso | **GR** | Greece | | | **TR** | Turkey |
| **BG** | Bulgaria | **HU** | Hungary | **ML** | Mali | **TT** | Trinidad and Tobago |
| **BJ** | Benin | **IE** | Ireland | **MN** | Mongolia | **UA** | Ukraine |
| **BR** | Brazil | **IL** | Israel | **MR** | Mauritania | **UG** | Uganda |
| **BY** | Belarus | **IS** | Iceland | **MW** | Malawi | **US** | United States of America |
| **CA** | Canada | **IT** | Italy | **MX** | Mexico | **UZ** | Uzbekistan |
| **CF** | Central African Republic | **JP** | Japan | **NE** | Niger | **VN** | Viet Nam |
| **CG** | Congo | **KE** | Kenya | **NL** | Netherlands | **YU** | Yugoslavia |
| **CH** | Switzerland | **KG** | Kyrgyzstan | **NO** | Norway | **ZW** | Zimbabwe |
| **CI** | Côte d'Ivoire | **KP** | Democratic People's Republic of Korea | **NZ** | New Zealand | | |
| **CM** | Cameroon | | | **PL** | Poland | | |
| **CN** | China | **KR** | Republic of Korea | **PT** | Portugal | | |
| **CU** | Cuba | **KZ** | Kazakstan | **RO** | Romania | | |
| **CZ** | Czech Republic | **LC** | Saint Lucia | **RU** | Russian Federation | | |
| **DE** | Germany | **LI** | Liechtenstein | **SD** | Sudan | | |
| **DK** | Denmark | **LK** | Sri Lanka | **SE** | Sweden | | |
| **EE** | Estonia | **LR** | Liberia | **SG** | Singapore | | |

# PROTEIN SCAFFOLDS FOR ANTIBODY MIMICS AND OTHER BINDING PROTEINS

## Background of the Invention

This invention relates to protein scaffolds useful, for example, for the generation of products having novel binding characteristics.

Proteins having relatively defined three-dimensional structures, commonly referred to as protein scaffolds, may be used as reagents for the design of engineered products. These scaffolds typically contain one or more regions which are amenable to specific or random sequence variation, and such sequence randomization is often carried out to produce libraries of proteins from which desired products may be selected. One particular area in which such scaffolds are useful is the field of antibody design.

A number of previous approaches to the manipulation of the mammalian immune system to obtain reagents or drugs have been attempted. These have included injecting animals with antigens of interest to obtain mixtures of polyclonal antibodies reactive against specific antigens, production of monoclonal antibodies in hybridoma cell culture (Koehler and Milstein, Nature 256:495, 1975), modification of existing monoclonal antibodies to obtain new or optimized recognition properties, creation of novel antibody fragments with desirable binding characteristics, and randomization of single chain antibodies (created by connecting the variable regions of the heavy and light chains of antibody molecules with a flexible peptide linker) followed by selection for antigen binding by phage display (Clackson et al., Nature 352:624, 1991).

In addition, several non-immunoglobulin protein scaffolds have been proposed for obtaining proteins with novel binding properties. For example, a

"minibody" scaffold, which is related to the immunoglobulin fold, has been designed by deleting three beta strands from a heavy chain variable domain of a monoclonal antibody (Tramontano et al., J. Mol. Recognit. 7:9, 1994). This protein includes 61 residues and can be used to present two hypervariable loops. These two loops have been randomized and products selected for antigen binding, but thus far the framework appears to have somewhat limited utility due to solubility problems. Another framework used to display loops has been tendamistat, a 74 residue, six-strand beta sheet sandwich held together by two disulfide bonds (McConnell and Hoess, J. Mol. Biol. 250:460, 1995). This scaffold includes three loops, but, to date, only two of these loops have been examined for randomization potential.

Other proteins have been tested as frameworks and have been used to display randomized residues on alpha helical surfaces (Nord et al., Nat. Biotechnol. 15:772, 1997; Nord et al., Protein Eng. 8:601, 1995), loops between alpha helices in alpha helix bundles (Ku and Schultz, Proc. Natl. Acad. Sci. USA 92:6552, 1995), and loops constrained by disulfide bridges, such as those of the small protease inhibitors (Markland et al., Biochemistry 35:8045, 1996; Markland et al., Biochemistry 35:8058, 1996; Rottgen and Collins, Gene 164:243, 1995; Wang et al., J. Biol. Chem. 270:12250, 1995).

## Summary of the Invention

The present invention provides a new family of proteins capable of evolving to bind any compound of interest. These proteins, which make use of a fibronectin or fibronectin-like scaffold, function in a manner characteristic of natural or engineered antibodies (that is, polyclonal, monoclonal, or single-chain antibodies) and, in addition, possess structural advantages. Specifically, the structure of these antibody mimics has been designed for optimal folding, stability, and solubility, even under conditions which normally

lead to the loss of structure and function in antibodies.

These antibody mimics may be utilized for the purpose of designing proteins which are capable of binding to virtually any compound (for example, any protein) of interest. In particular, the fibronectin-based molecules described herein may be used as scaffolds which are subjected to directed evolution designed to randomize one or more of the three fibronectin loops which are analogous to the complementarity-determining regions (CDRs) of an antibody variable region. Such a directed evolution approach results in the production of antibody-like molecules with high affinities for antigens of interest. In addition, the scaffolds described herein may be used to display defined exposed loops (for example, loops previously randomized and selected on the basis of antigen binding) in order to direct the evolution of molecules that bind to such introduced loops. A selection of this type may be carried out to identify recognition molecules for any individual CDR-like loop or, alternatively, for the recognition of two or all three CDR-like loops combined into a non-linear epitope.

Accordingly, the present invention features a protein that includes a fibronectin type III domain having at least one randomized loop, the protein being characterized by its ability to bind to a compound that is not bound by the corresponding naturally-occurring fibronectin.

In preferred embodiments, the fibronectin type III domain is a mammalian (for example, a human) fibronectin type III domain; and the protein includes the tenth module of the fibronectin type III ($^{10}$Fn3) domain. In such proteins, compound binding is preferably mediated by either one, two, or three $^{10}$Fn3 loops. In other preferred embodiments, the second loop of $^{10}$Fn3 may be extended in length relative to the naturally-occurring module, or the $^{10}$Fn3 may lack an integrin-binding motif. In these molecules, the integrin-binding motif may be replaced by an amino acid sequence in which a basic amino acid-

neutral amino acid-acidic amino acid sequence (in the N-terminal to C-terminal direction) replaces the integrin-binding motif; one preferred sequence is serine-glycine-glutamate. In another preferred embodiment, the fibronectin type III domain-containing proteins of the invention lack disulfide bonds.

Any of the fibronectin type II domain-containing proteins described herein may be formulated as part of a fusion protein (for example, a fusion protein which further includes an immunoglobulin $F_c$ domain, a complement protein, a toxin protein, or an albumin protein). In addition, any of the fibronectin type III domain proteins may be covalently bound to a nucleic acid (for example, an RNA), and the nucleic acid may encode the protein. Moreover, the protein may be a multimer, or, particularly if it lacks an integrin-binding motif, it may be formulated in a physiologically-acceptable carrier.

The present invention also includes features proteins that include a fibronectin type III domain having at least one mutation in a β-sheet sequence which changes the scaffold structure. Again, these proteins are characterized by their ability to bind to compound that are not bound by the corresponding naturally-occurring fibronectin.

In a related aspect, the invention further features nucleic acids encoding any of the proteins of the invention. In preferred embodiments, the nucleic acid is DNA or RNA.

In another related aspect, the invention also features a method for generating a protein which includes a fibronectin type III domain and which is pharmaceutically acceptable to a mammal, involving removing the integrin-binding domain of said fibronectin type III domain. This method may be applied to any of the fibronectin type III domain-containing proteins described above and is particularly useful for generating proteins for human therapeutic applications. The invention also features such fibronectin type III domain-containing proteins which lack integrin-binding domains.

In yet other related aspects, the invention features screening methods which may be used to obtain or evolve randomized fibronectin type III proteins capable of binding to compounds of interest, or to obtain or evolve compounds (for example, proteins) capable of binding to a particular protein containing a randomized fibronectin type III motif. In addition, the invention features screening procedures which combine these two methods, in any order, to obtain either compounds or proteins of interest.

In particular, the first screening method, useful for the isolation or identification of randomized proteins of interest, involves : (a) contacting the compound with a candidate protein, the candidate protein including a fibronectin type III domain having at least one randomized loop, the contacting being carried out under conditions that allow compound-protein complex formation; and (b) obtaining, from the complex, the protein which binds to the compound.

The second screening method, for isolating or identifying a compound which binds to a protein having a randomized fibronectin type III domain, involves: (a) contacting the protein with a candidate compound, the contacting being carried out under conditions that allow compound-protein complex formation; and (b) obtaining, from the complex, the compound which binds to the protein.

In preferred embodiments, the methods further involve either randomizing at least one loop of the fibronectin type III domain of the protein obtained in step (b) and repeating steps (a) and (b) using the further randomized protein, or modifying the compound obtained in step (b) and repeating steps (a) and (b) using the further modified compound. In addition, the compound is preferably a protein, and the fibronectin type III domain is preferably a mammalian (for example, a human) fibronectin type III domain. In other preferred embodiments, the protein includes the tenth module of the fibronectin

type III domain ($^{10}$Fn3), and binding is mediated by one, two or three $^{10}$Fn3 loops. In addition, the second loop of $^{10}$Fn3 may be extended in length relative to the naturally-occurring module, or $^{10}$Fn3 may lack an integrin-binding motif. Again, as described above, the integrin-binding motif may be replaced by an amino acid sequence in which a basic amino acid-neutral amino acid-acidic amino acid sequence (in the N-terminal to C-terminal direction) replaces the integrin-binding motif; one preferred sequence is serine-glycine-glutamate.

The selection methods described herein may be carried out using any fibronectin type III domain-containing protein. For example, the fibronectin type III domain-containing protein may lack disulfide bonds, or may be formulated as part of a fusion protein (for example, a fusion protein which further includes an immunoglobulin $F_c$ domain, a complement protein, a toxin protein, or an albumin protein). In addition, selections may be carried out using the fibronectin type III domain proteins covalently bound to nucleic acids (for example, RNAs or any nucleic acid which encodes the protein). Moreover, the selections may be carried out using fibronectin domain-containing protein multimers.

Preferably, the selections involve the immobilization of the binding target on a solid support. Preferred solid supports include columns (for example, affinity columns, such as agarose columns) or microchips.

As used herein, by "fibronectin type III domain" is meant a domain having 7 or 8 beta strands which are distributed between two beta sheets, which themselves pack against each other to form the core of the protein, and further containing loops which connect the beta strands to each other and are solvent exposed. There are at least three such loops at each edge of the beta sheet sandwich, where the edge is the boundary of the protein perpendicular to the direction of the beta strands. Preferably, a fibronectin type III domain includes a sequence which exhibits at least 30% amino acid identity, and preferably at

least 50% amino acid identity, to the sequence encoding the structure of the $^{10}$Fn3 domain referred to as "1ttg" (ID = "1ttg" (one ttg)) available from the Protein Data Base. Sequence identity referred to in this definition is determined by the Homology program, available from Molecular Simulation (San Diego, CA). The invention further includes polymers of $^{10}$Fn3-related molecules, which are an extension of the use of the monomer structure, whether or not the subunits of the polyprotein are identical or different in sequence.

By "naturally occurring fibronectin" is meant any fibronectin protein that is encoded by a living organism.

By "randomized" is meant including one or more amino acid alterations relative to a template sequence.

By a "protein" is meant any sequence of two or more amino acids, regardless of length, post-translation modification, or function. "Protein" and "peptide" are used interchangeably herein.

By "RNA" is meant a sequence of two or more covalently bonded, naturally occurring or modified ribonucleotides. One example of a modified RNA included within this term is phosphorothioate RNA.

By "DNA" is meant a sequence of two or more covalently bonded, naturally occurring or modified deoxyribonucleotides.

By a "nucleic acid" is meant any two or more covalently bonded nucleotides or nucleotide analogs or derivatives. As used herein, this term includes, without limitation, DNA, RNA, and PNA.

By "pharmaceutically acceptable" is meant a compound or protein that may be administered to an animal (for example, a mammal) without significant adverse medical consequences.

By "physiologically acceptable carrier" is meant a carrier which does not have a significant detrimental impact on the treated host and which retains the therapeutic properties of the compound with which it is administered. One

exemplary physiologically acceptable carrier is physiological saline. Other physiologically acceptable carriers and their formulations are known to one skilled in the art and are described, for example, in Remington's Pharmaceutical Sciences, (18th edition), ed. A. Gennaro, 1990, Mack Publishing Company, Easton, PA, incorporated herein by reference.

By "selecting" is meant substantially partitioning a molecule from other molecules in a population. As used herein, a "selecting" step provides at least a 2-fold, preferably, a 30-fold, more preferably, a 100-fold, and, most preferably, a 1000-fold enrichment of a desired molecule relative to undesired molecules in a population following the selection step. A selection step may be repeated any number of times, and different types of selection steps may be combined in a given approach.

By "binding partner," as used herein, is meant any molecule which has a specific, covalent or non-covalent affinity for a portion of a desired compound (for example, protein) of interest. Examples of binding partners include, without limitation, members of antigen/antibody pairs, protein/inhibitor pairs, receptor/ligand pairs (for example cell surface receptor/ligand pairs, such as hormone receptor/peptide hormone pairs), enzyme/substrate pairs (for example, kinase/substrate pairs), lectin/carbohydrate pairs, oligomeric or heterooligomeric protein aggregates, DNA binding protein/DNA binding site pairs, RNA/protein pairs, and nucleic acid duplexes, heteroduplexes, or ligated strands, as well as any molecule which is capable of forming one or more covalent or non-covalent bonds (for example, disulfide bonds) with any portion of another molecule (for example, a compound or protein).

By a "solid support" is meant, without limitation, any column (or column material), bead, test tube, microtiter dish, solid particle (for example, agarose or sepharose), microchip (for example, silicon, silicon-glass, or gold

chip), or membrane (for example, the membrane of a liposome or vesicle) to which an affinity complex may be bound, either directly or indirectly (for example, through other binding partner intermediates such as other antibodies or Protein A), or in which an affinity complex may be embedded (for example, through a receptor or channel).

The present invention provides a number of advantages. For example, as described in more detail below, the present antibody mimics exhibit improved biophysical properties, such as stability under reducing conditions and solubility at high concentrations. In addition, these molecules may be readily expressed and folded in prokaryotic systems, such as E. coli, in eukaryotic systems, such as yeast, and in in vitro translation systems, such as the rabbit reticulocyte lysate system. Moreover, these molecules are extremely amenable to affinity maturation techniques involving multiple cycles of selection, including in vitro selection using RNA-protein fusion technology (Roberts and Szostak, Proc. Natl. Acad. Sci USA 94:12297, 1997; Szostak et al., U.S.S.N. 09/007,005 and U.S.S.N. 09/247,190; Szostak et al. WO98/31700), phage display (see, for example, Smith and Petrenko, Chem. Rev. 97:317, 1997), and yeast display systems (see, for example, Boder and Wittrup, Nature Biotech. 15:553, 1997).

Other features and advantages of the present invention will be apparent from the following detailed description thereof, and from the claims.

## Brief Description of the Drawings

FIGURE 1 is a photograph showing a comparison between the structures of antibody heavy chain variable regions from camel (dark blue) and llama (light blue), in each of two orientations.

FIGURE 2 is a photograph showing a comparison between the

structures of the camel antibody heavy chain variable region (dark blue), the llama antibody heavy chain variable region (light blue), and a fibronectin type III module number 10 ($^{10}$Fn3) (yellow).

FIGURE 3 is a photograph showing a fibronectin type III module number 10 ($^{10}$Fn3), with the loops corresponding to the antigen-binding loops in IgG heavy chains highlighted in red.

FIGURE 4 is a graph illustrating a sequence alignment between a fibronectin type III protein domain and related protein domains.

FIGURE 5 is a photograph showing the structural similarities between a $^{10}$Fn3 domain and 15 related proteins, including fibronectins, tenascins, collagens, and undulin. In this photograph, the regions are labeled as follows: constant, dark blue; conserved, light blue; neutral, white; variable, red; and RGB integrin-binding motif (variable), yellow.

FIGURE 6 is a photograph showing space filling models of fibronectin III modules 9 and 10, in each of two different orientations. The two modules and the integrin binding loop (RGB) are labeled. In this figure, blue indicates positively charged residues, red indicates negatively charged residues, and white indicates uncharged residues.

FIGURE 7 is a photograph showing space filling models of fibronectin III modules 7-10, in each of three different orientiations. The four modules are labeled. In this figure, blue indicates positively charged residues, red indicates negatively charged residues, and white indicates uncharged residues.

FIGURE 8 is a photograph illustrating the formation, under different salt conditions, of RNA-protein fusions which include fibronectin type III domains.

FIGURE 9 is a series of photographs illustrating the selection of fibronectin type III domain-containing RNA-protein fusions, as measured by

PCR signal analysis.

FIGURE 10 is a graph illustrating an increase in the percent TNF-α binding during the selections described herein, as well as a comparison between RNA-protein fusion and free protein selections.

FIGURE 11 is a series of schematic representations showing IgG, $^{10}$Fn3, Fn-$CH_1$-$CH_2$-$CH_3$, and Fn-$CH_2$-$CH_3$ (clockwise from top left).

FIGURE 12 is a photograph showing a molecular model of Fn-$CH_1$-$CH_2$-$CH_3$ based on known three-dimensional structures of IgG (X-ray crystallography) and $^{10}$Fn3 (NMR and X-ray crystallography).

## Detailed Description

The novel antibody mimics described herein have been designed to be superior both to antibody-derived fragments and to non-antibody frameworks, for example, those frameworks described above.

The major advantage of these antibody mimics over antibody fragments is structural. These scaffolds are derived from whole, stable, and soluble structural modules found in human body fluid proteins. Consequently, they exhibit better folding and thermostability properties than antibody fragments, whose creation involves the removal of parts of the antibody native fold, often exposing amino acid residues that, in an intact antibody, would be buried in a hydrophobic environment, such as an interface between variable and constant domains. Exposure of such hydrophobic residues to solvent increases the likelihood of aggregation.

In addition, the antibody mimics described herein have no disulfide bonds, which have been reported to retard or prevent proper folding of antibody fragments under certain conditions. Since the present scaffolds do not rely on disulfides for native fold stability, they are stable under reducing conditions, unlike antibodies and their fragments which unravel upon disulfide bond

breakdown.

Moreover, these fibronectin-based scaffolds provide the functional advantages of antibody molecules. In particular, despite the fact that the $^{10}$Fn3 module is not an immunoglobulin, its overall fold is close to that of the variable region of the IgG heavy chain (Figure 2), making it possible to display the three fibronectin loops analogous to CDRs in relative orientations similar to those of native antibodies. Because of this structure, the present antibody mimics possess antigen binding properties that are similar in nature and affinity to those of antibodies, and a loop randomization and shuffling strategy may be employed in vitro that is similar to the process of affinity maturation of antibodies in vivo.

There are now described below exemplary fibronectin-based scaffolds and their use for identifying, selecting, and evolving novel binding proteins as well as their target ligands. These examples are provided for the purpose of illustrating, and not limiting, the invention.

$^{10}$Fn3 Structural Motif

The antibody mimics of the present invention are based on the structure of a fibronectin module of type III (Fn3), a common domain found in mammalian blood and structural proteins. This domain occurs more than 400 times in the protein sequence database and has been estimated to occur in 2% of the proteins sequenced to date, including fibronectins, tenscin, intracellular cytoskeletal proteins, and prokaryotic enzymes (Bork and Doolittle, Proc. Natl. Acad. Sci. USA 89:8990, 1992; Bork et al., Nature Biotech. 15:553, 1997; Meinke et al., J. Bacteriol. 175:1910, 1993; Watanabe et al., J. Biol. Chem. 265:15659, 1990). In particular, these scaffolds include, as templates, the tenth module of human Fn3 ($^{10}$Fn3), which comprises 94 amino acid residues. The overall fold of this domain is closely related to that of the smallest functional

antibody fragment, the variable region of the heavy chain, which comprises the entire antigen recognition unit in camel and llama IgG (Figure 1, 2). The major differences between camel and llama domains and the $^{10}$Fn3 domain are that (i) $^{10}$Fn3 has fewer beta strands (seven vs. nine) and (ii) the two beta sheets packed against each other are connected by a disulfide bridge in the camel and llama domains, but not in $^{10}$Fn3.

The three loops of $^{10}$Fn3 corresponding to the antigen-binding loops of the IgG heavy chain run between amino acid residues 21-31, 51-56, and 76-88 (Figure 3). The length of the first and the third loop, 11 and 12 residues, respectively, fall within the range of the corresponding antigen-recognition loops found in antibody heavy chains, that is, 10-12 and 3-25 residues, respectively. Accordingly, once randomized and selected for high antigen affinity, these two loops make contacts with antigens equivalent to the contacts of the corresponding loops in antibodies.

In contrast, the second loop of $^{10}$Fn3 is only 6 residues long, whereas the corresponding loop in antibody heavy chains ranges from 16-19 residues. To optimize antigen binding, therefore, the second loop of $^{10}$Fn3 is preferably extended by 10-13 residues (in addition to being randomized) to obtain the greatest possible flexibility and affinity in antigen binding. Indeed, in general, the lengths as well as the sequences of the CDR-like loops of the antibody mimics may be randomized during in vitro or in vivo affinity maturation (as described in more detail below).

The tenth human fibronectin type III domain, $^{10}$Fn3, refolds rapidly even at low temperature; its backbone conformation has been recovered within 1 second at 5°C. Thermodynamic stability of $^{10}$Fn3 is high ($\Delta G_U$ = 24 kJ/mol = 5.7 kcal/mol), correlating with its high melting temperature of 110°C.

One of the physiological roles of $^{10}$Fn3 is as a subunit of fibronectin, a glycoprotein that exists in a soluble form in body fluids and in an insoluble

form in the extracellular matrix (Dickinson et al., J. Mol. Biol. 236:1079, 1994). A fibronectin monomer of 220-250 kD contains 12 type I modules, two type II modules, and 17 fibronectin type III modules (Potts and Campbell, Curr. Opin.Cell Biol. 6:648, 1994). Different type III modules are involved in the binding of fibronectin to integrins, heparin, and chondroitin sulfate. $^{10}$Fn3 was found to mediate cell adhesion through an integrin-binding Arg-Gly-Asp (RGD) motif on one of its exposed loops. Similar RGD motifs have been shown to be involved in integrin binding by other proteins, such as fibrinogen, von Wellebrand factor, and vitronectin (Hynes et al., Cell 69:11, 1992). No other matrix- or cell-binding roles have been described for $^{10}$Fn3.

The observation that $^{10}$Fn3 has only slightly more adhesive activity than a short peptide containing RGD is consistent with the conclusion that the cell-binding activity of $^{10}$Fn3 is localized in the RGD peptide rather than distributed throughout the $^{10}$Fn3 structure (Baron et al., Biochemistry 31:2068, 1992). The fact that $^{10}$Fn3 without the RGD motif is unlikely to bind to other plasma proteins or extracellular matrix makes $^{10}$Fn3 a useful scaffold to replace antibodies. In addition, the presence of $^{10}$Fn3 in natural fibrinogen in the bloodstream suggests that $^{10}$Fn3 itself is unlikely to be immunogenic in the organism of origin.

In addition, we have determined that the $^{10}$Fn3 framework possesses exposed loop sequences tolerant of randomization, facilitating the generation of diverse pools of antibody mimics. This determination was made by examining the flexibility of the $^{10}$Fn3 sequence. In particular, the human $^{10}$Fn3 sequence was aligned with the sequences of fibronectins from other sources as well as sequences of related proteins (Figure 4), and the results of this alignment were mapped onto the three-dimensional structure of the human $^{10}$Fn3 domain (Figure 5). This alignment revealed that the majority of conserved residues are found in the core of the beta sheet sandwich, whereas the highly variable

residues are located along the edges of the beta sheets, including the N- and C-termini, on the solvent-accessible faces of both beta sheets, and on three solvent-accessible loops that serve as the hypervariable loops for affinity maturation of the antibody mimics. In view of these results, the randomization of these three loops are unlikely to have an adverse effect on the overall fold or stability of the $^{10}$Fn3 framework itself.

For the human $^{10}$Fn3 sequence, this analysis indicates that, at a minimum, amino acids 1-9, 44-50, 61-54, 82-94 (edges of beta sheets); 19, 21, 30-46 (even), 79-65 (odd) (solvent-accessible faces of both beta sheets); 21-31, 51-56, 76-88 (CDR-like solvent-accessible loops); and 14-16 and 36-45 (other solvent-accessible loops and beta turns) may be randomized to evolve new or improved compound-binding proteins. In addition, as discussed above, alterations in the lengths of one or more solvent exposed loops may also be included in such directed evolution methods. Alternatively, changes in the β-sheet sequences may also be used to evolve new proteins. These mutations change the scaffold and thereby indirectly alter loop structure(s). If this approach is taken, mutations should not saturate the sequence, but rather few mutations should be introduced. Preferably, no more than 10 amino acid changes, and, more preferably, no more than 3 amino acid changes should be introduced to the β-sheet sequences by this approach.

Fibronectin Fusions

The antibody mimics described herein may be fused to other protein domains. For example, these mimics may be integrated with the human immune response by fusing the constant region of an IgG ($F_c$) with a $^{10}$Fn3 module, preferably through the C-terminus of $^{10}$Fn3. The $F_c$ in such a $^{10}$Fn3-$F_c$ fusion molecule activates the complement component of the immune response and increases the therapeutic value of the antibody mimic. Similarly, a fusion

between $^{10}$Fn3 and a complement protein, such as C1q, may be used to target cells, and a fusion between $^{10}$Fn3 and a toxin may be used to specifically destroy cells that carry a particular antigen. In addition, $^{10}$Fn3 in any form may be fused with albumin to increase its half-life in the bloodstream and its tissue penetration. Any of these fusions may be generated by standard techniques, for example, by expression of the fusion protein from a recombinant fusion gene constructed using publically available gene sequences.

Fibronectin Scaffold Multimers

In addition to fibronectin monomers, any of the fibronectin constructs described herein may be generated as dimers or multimers of $^{10}$Fn3-based antibody mimics as a means to increase the valency and thus the avidity of antigen binding. Such multimers may be generated through covalent binding between individual $^{10}$Fn3 modules, for example, by imitating the natural $^{8}$Fn3-$^{9}$Fn3-$^{10}$Fn3 C-to-N-terminus binding or by imitating antibody dimers that are held together through their constant regions. A $^{10}$Fn3-Fc construct may be exploited to design dimers of the general scheme of $^{10}$Fn3-Fc::Fc-$^{10}$Fn3. The bonds engineered into the Fc::Fc interface may be covalent or non-covalent. In addition, dimerizing or multimerizing partners other than Fc can be used in $^{10}$Fn3 hybrids to create such higher order structures.

In particular examples, covalently bonded multimers may be generated by constructing fusion genes that encode the multimer or, alternatively, by engineering codons for cysteine residues into monomer sequences and allowing disulfide bond formation to occur between the expression products. Non-covalently bonded multimers may also be generated by a variety of techniques. These include the introduction, into monomer sequences, of codons corresponding to positively and/or negatively charged

residues and allowing interactions between these residues in the expression products (and therefore between the monomers) to occur. This approach may be simplified by taking advantage of charged residues naturally present in a monomer subunit, for example, the negatively charged residues of fibronectin. Another means for generating non-covalently bonded antibody mimics is to introduce, into the monomer gene (for example, at the amino- or carboxy-termini), the coding sequences for proteins or protein domains known to interact. Such proteins or protein domains include coil-coil motifs, leucine zipper motifs, and any of the numerous protein subunits (or fragments thereof) known to direct formation of dimers or higher order multimers.

Fibronectin-Like Molecules

Although $^{10}$Fn3 represents a preferred scaffold for the generation of antibody mimics, other molecules may be substituted for $^{10}$Fn3 in the molecules described herein. These include, without limitation, human fibronectin modules $^{1}$Fn3-$^{9}$Fn3 and $^{11}$Fn3-$^{17}$Fn3 as well as related Fn3 modules from non-human animals and prokaryotes. In addition, Fn3 modules from other proteins with sequence homology to $^{10}$Fn3, such as tenascins and undulins, may also be used. Modules from different organisms and parent proteins may be most appropriate for different applications; for example, in designing an antibody mimic, it may be most desirable to generate that protein from a fibronectin or fibronectin-like molecule native to the organism for which a therapeutic or diagnostic molecule is intended.

Directed Evolution of Scaffold-Based Binding Proteins

The antibody mimics described herein may be used in any technique for evolving new or improved binding proteins. In one particular example, the target of binding is immobilized on a solid support, such as a column resin or

microtiter plate well, and the target contacted with a library of candidate scaffold-based binding proteins. Such a library may consist of $^{10}$Fn3 clones constructed from the wild type $^{10}$Fn3 scaffold through randomization of the sequence and/or the length of the $^{10}$Fn3 CDR-like loops. If desired, this library may be an RNA-protein fusion library generated, for example, by the techniques described in Szostak et al., U.S.S.N. 09/007,005 and 09/247,190; Szostak et al., WO98/31700; and Roberts & Szostak, Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 12297-12302. Alternatively, it may be a DNA-protein library (for example, as described in Lohse, DNA-Protein Fusions and Uses Thereof, U.S.S.N. 60/110,549, filed December 2, 1998 and _____, filed December 2, 1999). The fusion library is incubated with the immobilized target, the support is washed to remove non-specific binders, and the tightest binders are eluted under very stringent conditions and subjected to PCR to recover the sequence information or to create a new library of binders which may be used to repeat the selection process, with or without further mutagenesis of the sequence. A number of rounds of selection may be performed until binders of sufficient affinity for the antigen are obtained.

In one particular example, the $^{10}$Fn3 scaffold may be used as the selection target. For example, if a protein is required that binds a specific peptide sequence presented in a ten residue loop, a single $^{10}$Fn3 clone is constructed in which one of its loops has been set to the length of ten and to the desired sequence. The new clone is expressed *in vivo* and purified, and then immobilized on a solid support. An RNA-protein fusion library based on an appropriate scaffold is then allowed to interact with the support, which is then washed, and desired molecules eluted and re-selected as described above.

Similarly, the $^{10}$Fn3 scaffold may be used to find natural proteins that interact with the peptide sequence displayed in a $^{10}$Fn3 loop. The $^{10}$Fn3 protein is immobilized as described above, and an RNA-protein fusion library is

screened for binders to the displayed loop. The binders are enriched through multiple rounds of selection and identified by DNA sequencing.

In addition, in the above approaches, although RNA-protein libraries represent exemplary libraries for directed evolution, any type of scaffold-based library may be used in the selection methods of the invention.

Use

The antibody mimics described herein may be evolved to bind any antigen of interest. These proteins have thermodynamic properties superior to those of natural antibodies and can be evolved rapidly in vitro. Accordingly, these antibody mimics may be employed in place of antibodies in all areas in which antibodies are used, including in the research, therapeutic, and diagnostic fields. In addition, because these scaffolds possess solubility and stability properties superior to antibodies, the antibody mimics described herein may also be used under conditions which would destroy or inactivate antibody molecules. Finally, because the scaffolds of the present invention may be evolved to bind virtually any compound, these molecules provide completely novel binding proteins which also find use in the research, diagnostic, and therapeutic areas.

Experimental Results

Exemplary scaffold molecules described above were generated and tested, for example, in selection protocols, as follows.

Library construction

A complex library was constructed from three fragments, each of which contained one randomized area corresponding to a CDR-like loop. The fragments were named BC, DE, and FG, based on the names of the

CDR-H-like loops contained within them; in addition to $^{10}$Fn3 and a randomized sequence, each of the fragments contained stretches encoding an N-terminal $His_6$ domain or a C-terminal FLAG peptide tag. At each junction between two fragments (i.e., between the BC and DE fragments or between the DE and FG fragments), each DNA fragment contained recognition sequences for the EarI Type IIS restriction endonuclease. This restriction enzyme allowed the splicing together of adjacent fragments while removing all foreign, non-$^{10}$Fn3, sequences. It also allows for a recombination-like mixing of the three $^{10}$Fn3 fragments between cycles of mutagenesis and selection.

Each fragment was assembled from two overlapping oligonucleotides, which were first annealed, then extended to form the double-stranded DNA form of the fragment. The oligonucleotides that were used to construct and process the three fragments are listed below; the "Top" and "Bottom" species for each fragment are the oligonucleotides that contained the entire $^{10}$Fn3 encoding sequence. In these oligonucleotides designations, "N" indicates A, T, C, or G; and "S" indicates C or G.

**HfnLbcTop (His):**

5'- GG AAT TCC TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA TTT ACA ATT ACA ATG CAT CAC CAT CAC CAT CAC GTT TCT GAT GTT CCG AGG GAC CTG GAA GTT GTT GCT GCG ACC CCC ACC AGC-3' (SEQ ID NO: 1)

**HfnLbcTop (an alternative N-terminus):**

5'- GG AAT TCC TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA TTT ACA ATT ACA ATG GTT TCT GAT GTT CCG AGG GAC CTG GAA GTT GTT GCT GCG ACC CCC ACC AGC-3' (SEQ ID NO: 2)

**HFnLBCBot-flag8:**

5'-AGC GGA TGC CTT GTC GTC GTC GTC CTT GTA GTC GCT CTT CCC TGT TTC TCC GTA AGT GAT CCT GTA ATA TCT (SNN)7 CCA GCT GAT CAG TAG GCT GGT GGG GGT CGC AGC -3' (SEQ ID NO: 3)

**HFnBC3'-flag8:**

5'-AGC GGA TGC CTT GTC GTC GTC GTC CTT GTA GTC GCT CTT CCC TGT TTC TCC GTA AGT GAT CC-3' (SEQ ID NO: 4)

**HFnLDETop:**

5'- GG AAT TCC TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA TTT ACA ATT ACA ATG CAT CAC CAT CAC CAT CAC CTC TTC ACA GGA GGA AAT AGC CCT GTC C-3' (SEQ ID NO: 5)

**HFnLDEBot-flag8:**

5'-AGC GGA TGC CTT GTC GTC GTC GTC CTT GTA GTC GCT CTT CGT ATA ATC AAC TCC AGG TTT AAG GCC GCT GAT GGT AGC TGT (SNN)4 AGG CAC AGT GAA CTC CTG GAC AGG GCT ATT TCC TCC TGT -3' (SEQ ID NO: 6)

**HFnDE3'-flag8:**

5'-AGC GGA TGC CTT GTC GTC GTC GTC CTT GTA GTC GCT CTT CGT ATA ATC AAC TCC AGG TTT AAG G-3' (SEQ ID NO: 7)

**HFnLFGTop:**

5'- GG AAT TCC TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA TTT ACA ATT ACA ATG CAT CAC CAT CAC CAT CAC CTC TTC TAT ACC ATC ACT GTG TAT GCT GTC-3' (SEQ ID NO: 8)

**HFnLFGBot-flag8:**

5'-AGC GGA TGC CTT GTC GTC GTC GTC CTT GTA GTC TGT TCG GTA ATT AAT GGA AAT TGG (SNN)10 AGT GAC AGC ATA CAC AGT GAT GGT ATA -3' (SEQ ID NO: 9)

**HFnFG3'-flag8:**

5'-AGC GGA TGC CTT GTC GTC GTC GTC CTT GTA GTC TGT TCG GTA ATT AAT GGA AAT TGG -3' (SEQ ID NO: 10)

**T7Tmv (introduces T7 promoter and TMV untranslated region needed for in vitro translation):**

5'- GCG TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA TTT ACA ATT ACA-3' (SEQ ID NO: 11)

**ASAflag8:**

5'-AGC GGA TGC CTT GTC GTC GTC GTC CTT GTA GTC-3' (SEQ ID NO: 12)

**Unispl-s (spint oligonucleotide used to ligate mRNA to the puromycin-containing linker, described by Roberts et al, 1997, supra):**

5'-TTTTTTTTTNAGCGGATGC-3' (SEQ ID NO: 13)

**A18—2PEG (DNA-puromycin linker):**

5'-(A)18(PEG)2CCPur (SEQ ID NO: 14)

The pairs of oligonucleotides (500 pmol of each) were annealed in 100 µL of 10 mM Tris 7.5, 50 mM NaCl for 10 minutes at 85°C, followed by a slow (0.5-1 hour) cooling to room temperature. The annealed fragments with

single-stranded overhangs were then extended using 100 U Klenow (New England Biolabs, Beverly, MA) for each 100 µL aliquot of annealed oligos, and the buffer made of 838.5 µl $H_2O$, 9 µl 1 M Tris 7.5, 5 µl 1M $MgCl_2$, 20 µl 10 mM dNTPs, and 7.5 µl 1M DTT. The extension reactions proceeded for 1 hour at 25°C.

Next, each of the double-stranded fragments was transformed into a RNA-protein fusion (PROfusion™) using the technique developed by Szostak et al., U.S.S.N. 09/007,005 and U.S.S.N. 09/247,190; Szostak et al., WO98/31700; and Roberts & Szostak, Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 12297-12302. Briefly, the fragments were transcribed using an Ambion in vitro transcription kit, MEGAshortscript (Ambion, Austin, TX), and the resulting mRNA was gel-purified and ligated to a DNA-puromycin linker using DNA ligase. The mRNA-DNA-puromycin molecule was then translated using the Ambion rabbit reticulocyte lysate-based translation kit. The resulting mRNA-DNA-puromycin-protein PROfusion™ was purified using Oligo(dT) cellulose, and a complementary DNA strand was synthesized using reverse transcriptase and the RT primers described above (Unisplint-S or flagASA), following the manufacturer's instructions.

The PROfusion™ obtained for each fragment was next purified on the resin appropriate to its peptide purification tag, i.e., on Ni-NTA agarose for the $His_6$-tag and M2 agarose for the FLAG-tag, following the procedure recommended by the manufacturer. The DNA component of the tag-binding PROfusions™ was amplified by PCR using Pharmacia Ready-to-Go PCR Beads, 10 pmol of 5' and 3' PCR primers, and the following PCR program (Pharmacia, Piscataway, NJ): Step 1: 95°C for 3 minutes; Step 2: 95°C for 30 seconds, 58/62°C for 30 seconds, 72°C for 1 minute, 20/25/30 cycles, as required; Step 3: 72°C for 5 minutes; Step 4: 4°C until end.

The resulting DNA was cleaved by 5 U EarI (New England Biolabs)

per1 ug DNA; the reaction took place in T4 DNA Ligase Buffer (New England Biolabs) at 37°C, for 1 hour, and was followed by an incubation at 70°C for15 minutes to inactivate Ear I. Equal amounts of the BC, DE, and FG fragments were combined and ligated to form a full-length $^{10}$Fn3 gene with randomized loops. The ligation required 10 U of fresh EarI (New England Biolabs) and 20 U of T4 DNA Ligase (Promega, Madison, WI), and took 1 hour at 37°C.

Three different libraries were made in the manner described above. Each contained the form of the FG loop with 10 randomized residues. The BC and the DE loops of the first library bore the wild type $^{10}$Fn3 sequence; a BC loop with 7 randomized residues and a wild type DE loop made up the second library; and a BC loop with 7 randomized residues and a DE loop with 4 randomized residues made up the third library. The complexity of the FG loop in each of these three libraries was $10^{13}$; the further two randomized loops provided the potential for a complexity too large to be sampled in a laboratory.

The three libraries constructed were combined into one master library in order to simplify the selection process; target binding itself was expected to select the most suitable library for a particular challenge. PROfusions™ were obtained from the master library following the general procedure described in Szostak et al., U.S.S.N. 09/007,005 and 09/247,190; Szostak et al., WO98/31700; and Roberts & Szostak, Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 12297-12302 (Figure 8).

Fusion Selections

The master library in the PROfusion™ form was subjected to selection for binding to TNF-α. Two protocols were employed: one in which the target was immobilized on an agarose column and one in which the target was immobilized on a BIACORE chip. First, an extensive optimization of conditions to minimize background binders to the agarose column yielded the

favorable buffer conditions of 50 mM HEPES pH 7.4, 0.02% Triton, 100 µg/ml Sheared Salmon Sperm DNA. In this buffer, the non-specific binding of the $^{10}$Fn3 RNA fusion to TNF-α Sepharose was 0.3%. The non-specific binding background of the $^{10}$Fn3 RNA-DNA to TNF-α Sepharose was found to be 0.1%.

During each round of selection on TNF-α Sepharose, the Profusion™ library was first preincubated for an hour with underivatized Sepharose to remove any remaining non-specific binders; the flow-through from this pre-clearing was incubated for another hour with TNF-α Sepharose. The TNF-α Sepharose was washed for 3-30 minutes.

After each selection, the PROfusion™ DNA that had been eluted from the solid support with 0.3 M NaOH or 0.1M KOH was amplified by PCR; a DNA band of the expected size persisted through multiple rounds of selection (Figure 9); similar results were observed in the two alternative selection protocols, and only the data from the agarose column selection is shown in Figure 9.

In the first seven rounds, the binding of library PROfusions™ to the target remained low; in contrast, when free protein was translated from DNA pools at different stages of the selection, the proportion of the column binding species increased significantly between rounds (Figure 10). Similar selections may be carried out with any other binding species target (for example, IL-1 and IL-13).

Animal Studies

Wild-type $^{10}$Fn3 contains an integrin-binding tripepetide motif, Arginine 78 - Glycine 79 - Aspartate 80 (the "RGD motif') at the tip of the FG loop. In order to avoid integrin binding and a potential inflammatory response based on this tripeptide in vivo, a mutant form of $^{10}$Fn3 was generated that

contained an inert sequence, Serine 78 - Glycine 79 - Glutamate 80 (the "SGE mutant"), a sequence which is found in the closely related, wild-type $^{11}$Fn3 domain. This SGE mutant was expressed as an N-terminally $His_6$-tagged, free protein in E. coli, and purified to homogeneity on a metal chelate column followed by a size exclusion column.

In particular, the DNA sequence encoding $His_6$-$^{10}$Fn3(SGE) was cloned into the pET9a expression vector and transformed into BL21 DE3 pLysS cells. The culture was then grown in LB broth containing 50 µg/mL kanamycin at 37°C, with shaking, to $A_{560}$=1.0, and was then induced with 0.4 mM IPTG. The induced culture was further incubated, under the same conditions, overnight (14-18 hours); the bacteria were recovered by standard, low speed centrifugation. The cell pellet was resuspended in 1/50 of the original culture volume of lysis buffer (50 mM Tris 8.0, 0.5 M NaCl, 5% glycerol, 0.05% Triton X-100, and 1 mM PMSF), and the cells were lysed by passing the resulting paste through a Microfluidics Corporation Microfluidizer M110-EH, three times. The lysate was clarified by centrifugation, and the supernatant was filtered through a 0.45 µm filter followed by filtration through a 0.2 µm filter. 100 mL of the clarified lysate was loaded onto a 5 mL Talon cobalt column (Clontech, Palo Alto, CA), washed by 70 mL of lysis buffer, and eluted with a linear gradient of 0-30 mM imidazole in lysis buffer. The flow rate through the column through all the steps was 1 mL/min. The eluted protein was concentrated 10-fold by dialysis (MW cutoff = 3,500) against 15,000-20,000 PEG. The resulting sample was dialysed into buffer 1 (lysis buffer without the glycerol), then loaded, 5 mL at a time, onto a 16 x 60 mm Sephacryl 100 size exclusion column equilibrated in buffer 1. The column was run at 0.8 mL/min, in buffer 1; all fractions that contained a protein of the expected MW were pooled, concentrated 10X as described above, then dialyzed into PBS. Toxikon (MA) was engaged to perform endotoxin screens

and animal studies on the resulting sample.

In these animal studies, the endotoxin levels in the samples examined to date have been below the detection level of the assay. In a preliminary toxicology study, this protein was injected into two mice at the estimated 100X therapeutic dose of 2.6 mg/mouse. The animals survived the two weeks of the study with no apparent ill effects. These results suggest that $^{10}$Fn3 may be incorporated safely into an IV drug.

Alternative Constructs for In Vivo Use

To extend the half life of the 8 kD $^{10}$Fn3 domain, a larger molecule has also been constructed that mimics natural antibodies. This $^{10}$Fn3-$F_c$ molecule contains the -$CH_1$-$CH_2$-$CH_3$ (Figure 11) or -$CH_2$-$CH_3$ domains of the IgG constant region of the host; in these constructs, the $^{10}$Fn3 domain is grafted onto the N-terminus in place of the IgG $V_H$ domain (Figures 11 and 12). Such antibody-like constructs are expected to improve the pharmacokinetics of the protein as well as its ability to harness the natural immune response.

In order to construct the murine form of the $^{10}$Fn3-$CH_1$-$CH_2$-$CH_3$ clone, the -$CH_1$-$CH_2$-$CH_3$ region was first amplified from a mouse liver spleen cDNA library (Clontech), then ligated into the pET25b vector. The primers used in the cloning were 5' Fc Nest and 3' 5 Fc Nest, and the primers used to graft the appropriate restriction sites onto the ends of the recovered insert were 5' Fc HIII and 3' Fc Nhe:

5' Fc Nest 5'GCG GCA GGG TTT GCT TAC TGG GGC CAA GGG 3' (SEQ ID NO: 15);

3' Fc Nest 5'GGG AGG GGT GGA GGT AGG TCA CAG TCC 3' (SEQ ID NO: 16);

3' Fc Nhe 5' TTT GCT AGC TTT ACC AGG AGA GTG GGA GGC 3' (SEQ

ID NO: 17); and

5' Fc HIII 5' AAA AAG CTT GCC AAA ACG ACA CCC CCA TCT GTC 3' (SEQ ID NO: 18).

Further PCR is used to remove the $CH_1$ region from this clone and create the Fc part of the shorter, $^{10}$Fn3-$CH_2$-$CH_3$ clone. The sequence encoding $^{10}$Fn3 is spliced onto the 5' end of each clone; either the wild type $^{10}$Fn3 cloned from the same mouse spleen cDNA !ibrary or a modified $^{10}$Fn3 obtained by mutagenesis or randomization of the molecules can be used. The oligonucleotides used in the cloning of murine wild-type $^{10}$Fn3 were:

Mo 5PCR-NdeI:

5' CATATGGTTTCTGATATTCCGAGAGATCTGGAG 3' (SEQ ID NO: 19);

Mo5PCR-His-NdeI (for an alternative N-terminus with the $His_6$ purification tag):

5' CAT ATG CAT CAC CAT CAC CAT CAC GTT TCT GAT ATT CCG AGA G 3' (SEQ ID NO: 20); and

Mo3PCR-EcoRI: 5' GAATTCCTATGTTTTATAATTGATGGAAAC3' (SEQ ID NO: 21).

The human equivalents of the clones are constructed using the same strategy with human oligonucleotide sequences.

Other embodiments are within the claims.

All publications, patents, and patent applications mentioned herein are hereby incorporated by reference.

What is claimed is:

Claims

1. A protein comprising a fibronectin type III domain having at least one randomized loop, said protein being characterized by its ability to bind to a compound that is not bound by the corresponding naturally-occurring fibronectin.

2. The protein of claim 1, wherein said fibronectin type III domain is a mammalian fibronectin type III domain.

3. The protein of claim 2, wherein said fibronectin type III domain is a human fibronectin type III domain.

4. The protein of claim 1, wherein said protein comprises the tenth module of said fibronectin type III domain ($^{10}$Fn3).

5. The protein of claim 4, wherein said compound binding is mediated by one $^{10}$Fn3 loop.

6. The protein of claim 4, wherein said compound binding is mediated by two $^{10}$Fn3 loops.

7. The protein of claim 4, wherein said compound binding is mediated by three $^{10}$Fn3 loops.

8. The protein of claim 4, wherein the second loop of said $^{10}$Fn3 is extended in length relative to the naturally-occurring module.

9. The protein of claim 4, wherein said $^{10}$Fn3 lacks an integrin-binding motif.

10. The protein of claim 9, wherein said integrin-binding motif is replaced by an amino acid sequence comprising a basic amino acid-neutral amino acid-acidic amino acid motif.

11. The protein of claim 10, wherein said integrin-binding motif is replaced by an amino acid sequence comprising serine-glycine-glutamate.

12. The protein of claim 1, wherein said protein lacks disulfide bonds.

13. The protein of claim 1, wherein said protein is part of a fusion protein.

14. The protein of claim 13, wherein said fusion protein further comprises an immunoglobulin $F_c$ domain.

15. The protein of claim 13, wherein said fusion protein further comprises a complement protein.

16. The protein of claim 13, wherein said fusion protein further comprises a toxin protein.

17. The protein of claim 13, wherein said fusion protein further comprises an albumin protein.

18. The protein of claim 1, wherein said protein is covalently bound to a nucleic acid.

19. The protein of claim 18, wherein said nucleic acid encodes said protein.

20. The protein of claim 18, wherein said nucleic acid is RNA.

21. The protein of claim 1, wherein said protein is a multimer.

22. The protein of claim 1 or 9, wherein said protein is formulated in a physiologically-acceptable carrier.

23. A nucleic acid encoding the protein of claim 1 or 4.

24. The nucleic acid of claim 23, wherein said nucleic acid is DNA.

25. The nucleic acid of claim 23, wherein said nucleic acid is RNA.

26. A method for generating a protein comprising a fibronectin type III domain which is pharmaceutically acceptable to a mammal, said method comprising removing an integrin-binding domain from said fibronectin type III domain.

27. The method of claim 26, wherein said integrin binding motif is replaced by an amino acid sequence comprising a basic amino acid-neutral amino acid-acidic amino acid motif.

28. The protein of claim 27, wherein said integrin-binding motif is replaced by an amino acid sequence comprising serine-glycine-glutamate.

29. The method of claim 26, wherein said at least one loop of said fibronectin type III domain is randomized.

30. The method of claim 26, wherein said protein comprises the tenth module of said fibronectin type III domain.

31. The protein of claim 26, wherein said protein is part of a fusion protein.

32. The protein of claim 31, wherein said fusion protein further comprises an immunoglobulin $F_c$ domain.

33. The protein of claim 31, wherein said fusion protein further comprises a complement protein.

34. The protein of claim 31, wherein said fusion protein further comprises a toxin protein.

35. The protein of claim 31, wherein said fusion protein further comprises an albumin protein.

36. The method of claim 26, wherein said mammal is a human.

37. A method for obtaining a protein which binds to a compound, said method comprising:

(a) contacting said compound with a candidate protein, said candidate protein comprising a fibronectin type III domain having at least one randomized loop, said contacting being carried out under conditions that allow compound-protein complex formation; and

(b) obtaining, from said complex, said protein which binds to said compound.

38. A method for obtaining a compound which binds to a protein, said protein comprising a fibronectin type III domain having at least one randomized loop, said method comprising:

(a) contacting said protein with a candidate compound, said contacting being carried out under conditions that allow compound-protein complex formation; and

(b) obtaining, from said complex, said compound which binds to said protein.

39. The method of claim 37, said method further comprising randomizing at least one loop of said fibronectin type III domain of said protein obtained in step (b) and repeating said steps (a) and (b) using said further randomized protein.

40. The method of claim 38, said method further comprising modifying said compound obtained in step (b) and repeating said steps (a) and (b) using said further modified compound.

41. The method of claim 37 or 38, wherein said compound is a protein.

42. The method of claim 37 or 38, wherein said fibronectin type III domain is a mammalian fibronectin type III domain.

43. The method of claim 42, wherein said fibronectin type III domain is a human fibronectin type III domain.

44. The method of claim 37 or 38, wherein said protein comprises the tenth module of said fibronectin type III domain ($^{10}$Fn3).

45. The method of claim 44, wherein said compound binding is mediated by one $^{10}$Fn3 loop.

46. The method of claim 44, wherein said compound binding is mediated by two $^{10}$Fn3 loops.

47. The method of claim 44, wherein said compound binding is mediated by three $^{10}$Fn3 loops.

48. The method of claim 44, wherein the second loop of said $^{10}$Fn3 is extended in length relative to the naturally-occurring module.

49. The method of claim 44, wherein said $^{10}$Fn3 lacks an integrin-binding motif.

50. The method of claim 37, wherein said compound is immobilized on a solid support.

51. The method of claim 38, wherein said protein is immobilized on

a solid support.

52. The method of claim 50 or 51, wherein said solid support is a column or microchip.

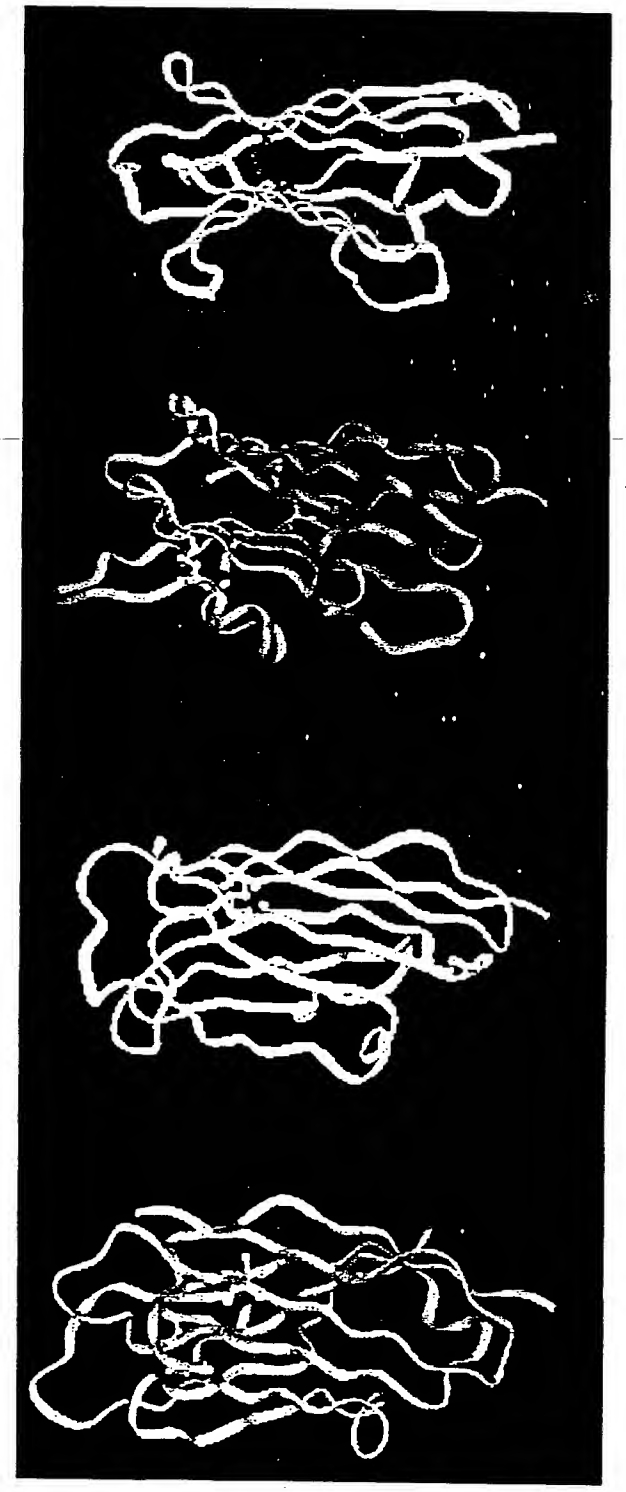


Fig. 1

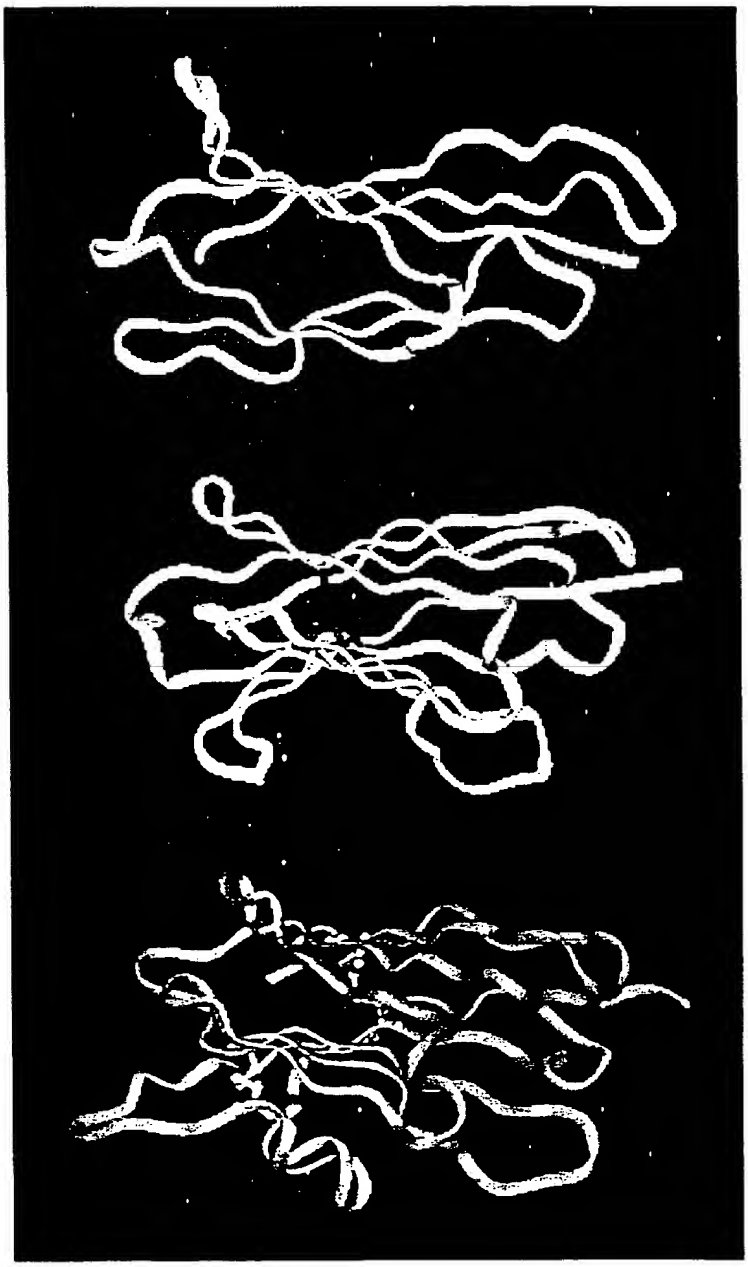
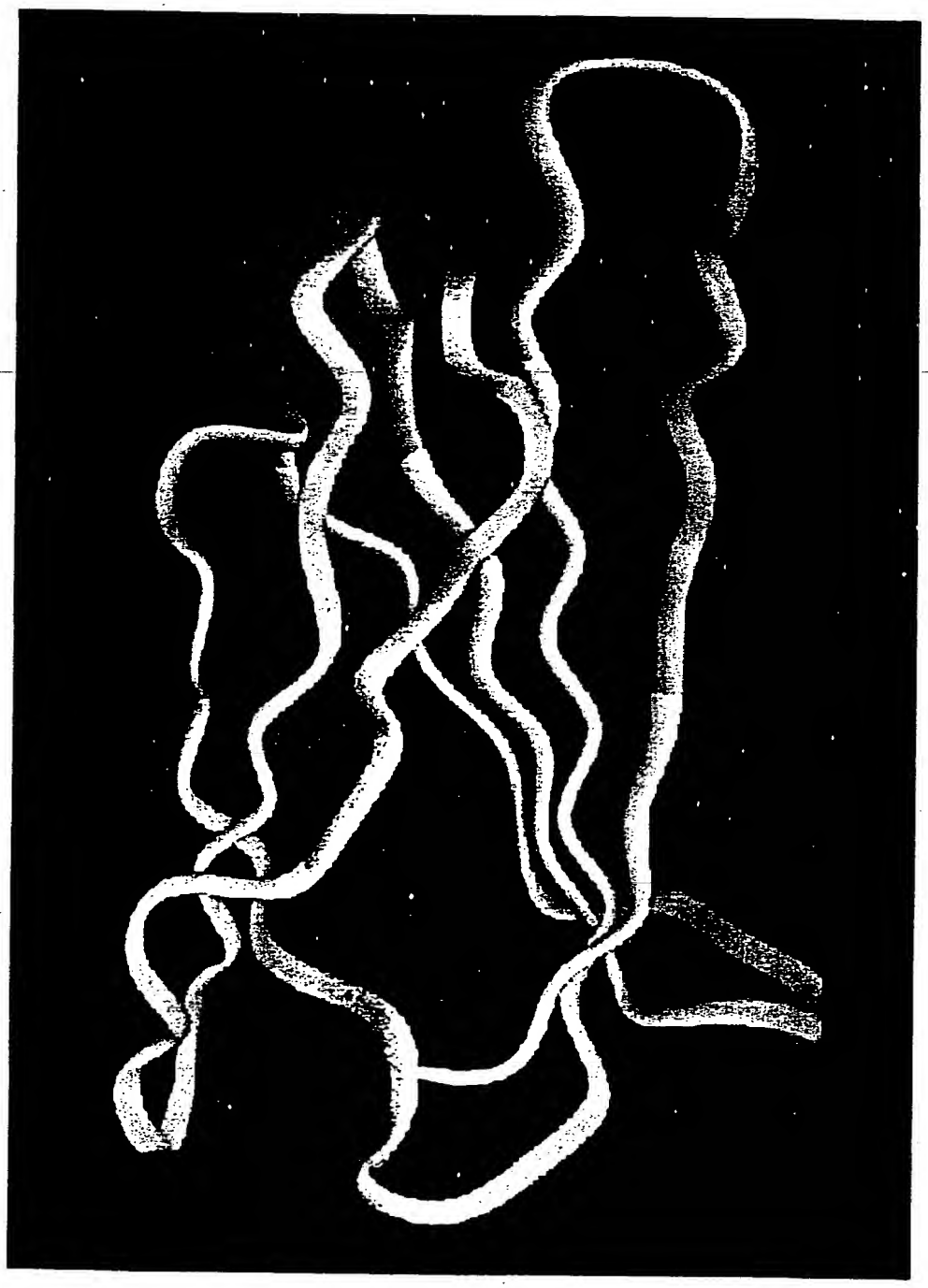


Fig. 2



F. g. 3

```
                1        9 10      19 20      29 30    37   38      47 48      57 58      67 68      77 78      87 88   94
Hs FND  VSDVPRD-LE VVAATPTSLL ISWDAPAVTV RYYRITYG-- ETGGNSPVQE FTVPGSKSTA TISGLKPGVD YTITVYAVTG RGDSPASSKP ISINYRT      94
Bt FN   VSDVPRD-LE VIAATPTSLL ISWDAPAVTV RYYRITYG-- ETGGSSPVQE FTVPGSKSTA TISGLKPGVD YTITVYAVTG RGDSPASSKP VSINYRT    1510
Rn FN   VSDVPRD-LE VIASTPTSLL ISWEpPAVSV RYYRITYG-- ETGGNSPVQE FTVPGSKSTA TINnIKPGAD YTITLYAVTG RGDSPASSKP VSINYQT    1611
Mm FN   VSDIPRD-LE VIASTPTSLL ISWEpPAVSV RYYRITYG-- ETGGNSPVQE FTVPGSKSTA TINnIKPGAD YTITLYAVTG RGDSPASSKP VSINYKT     712
Oc FN   VSDVPRD-LE VIASTPTSLL ISWEXPAVTV RYYRITYG-- ETppN----- ---------- ---------- ---------- ---------- -------      64
Gg FN   VSDVPRD-LE VNpTSPTSLE ISWDAPAVTV RYYRITYG-- ETGGSSPVQE FTVPGTMSrA TITGLKPGVD YTITVYAVTG RGDSPASSKP VTVTYKT     443
Xl FN   VSDVPTD-LE VTSSSPNTLT ISWEAPAVSV RYYRITYS-- QTGGhGPEKE FTVPGTSNTA TIRGLNPGVS YTITVYAVTG RGDSPASSKP LTIIHKT    1611
Cf FN   AiDAPSn-Lr FLATTPnSLL VSWQpPrArI TGYIIKYe-- kpGSpprEVV prprPGVTeA TITGLePGTE YTIQVIALKn NQKSepLIGr kKTdEL-     197
Ec FN   AiDAPSn-Lh FLATTPnSLL ISWQpPrArI TGYIIKYe-- kpGSpprEVV prphPGVTeA TITGLePGTE YTIQVIAIKn NQKSepLIGr rKTdEp-     197
Hs TC   VS-PPKD-Lv VTeVTeeTVN LAWDn-eMrV TeYLVVYTP- -THEGGlEMQ FrVPGDQTST IIQeLePGVE YFIRVFAILe NKKSipVSAr V------     686
Ss TP   VS-PPKD-Li VTeVTeeTVN LAWDn-eMrV TeYLIVYTP- -THEdGlEMQ FrVPGDQTST TIReLePGVE YFIRVFAILe NKKSipVSAr V------     686
Mm TX   MiDGPQD-Lr VVAVTPTTLD LSWlrPQAeV DrFVVSYV-- --SAGNqRVr LeVPPEADrT QLTdLMPGVE YVVTVTAERG HAVSypASIr ANTG---     889
Hs CAP  TlpVPvvSLn IYdVGPTTMH VQWQp-VGGA TGYILSYkPV kDTEpTrpKE VrLGPTVNdM QLTdLVPNTE YAVTVQAVLh dLTSepVTVr e------    1551
Oc C12  TlpVPvvSLN IYdVGPTTMH VQWQp-VGGA TGYILSYkPV kDTEpTrpKp QdVkLRdVTH ---------- ---------- ---------- -------     322
Gg C14  LalpmaSDLk LYdVShSSMR AKWnG-VAGA TGYMILYAPL TEGLAAdEKE IkIGEASTeL ELDGLLPNTE YTVTVYAMF- ---------- -------     508
Hs U1   LalpmaSDLl LYdVTenSMR VKWDA-VpGA SGYLILYAPL TEGLAGdEKE MkIGETHTdI ELSGLLPNTE YTVTVYAMFG eeASDpVTGq e------     321

var.    --- -- -  -  --       -----   -    -  ---------  - -  - -    - -            - --  --  - -  - -

cons.     P  L  V     SL   W     V   Y I Y                        I  L PGVD Y ITV A   G     S     P
          M     I     TV         A   F V                        L  I  NTE   VQL     N         R
                L      M         I     L                             AS    R     E
                 F
```

| | | | |
|---|---|---|---|
| CAP | Collagen alpha precursor | BOLD | identical to Hs FND |
| C12 | Collagen type 12 | | |
| FND | Fibronectin type III domain | lower case | non-conservative substitution (charge reversal, change between hydrophobic and charged, addition or removal of P) |
| FN | Fibronectin | | |
| TP | Tenascin precursor | - | position of non-conservative substitutions |
| TC | Tenascin-C | | |
| U1 | Undulin 1 | | |

| | | |
|---|---|---|
| Bt | Bovis taurus | cow |
| Cf | Canis familiaris | dog |
| Ec | Equus caballis | horse |
| Ss | Sus scrofa | pig |
| Hs | Homo sapiens | human |
| Oc | Oryctolagus cuniculus | rabbit |
| Xl | Xenupus laevis | African clawed frog |

Fig. 4

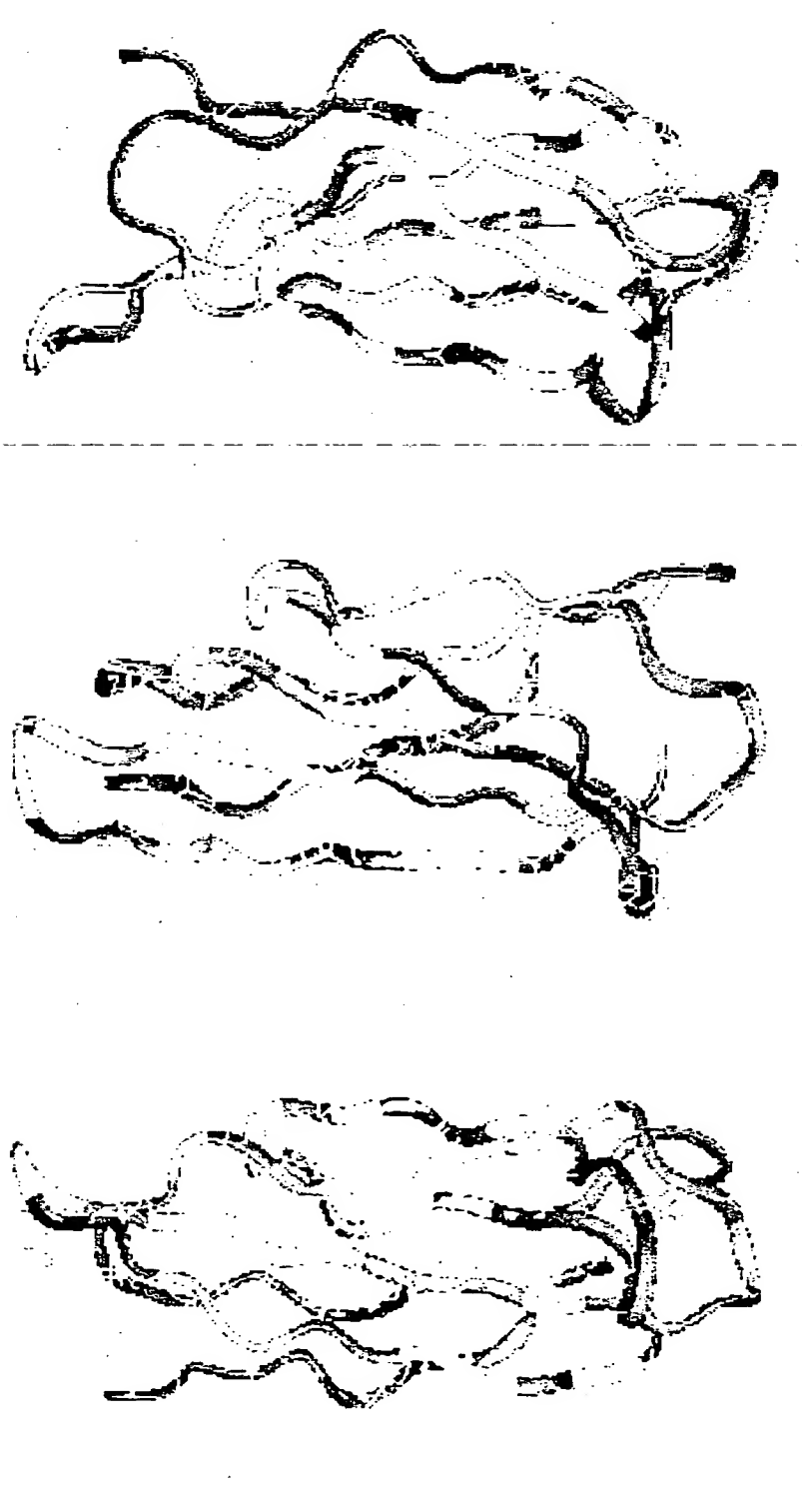


Fig. 5

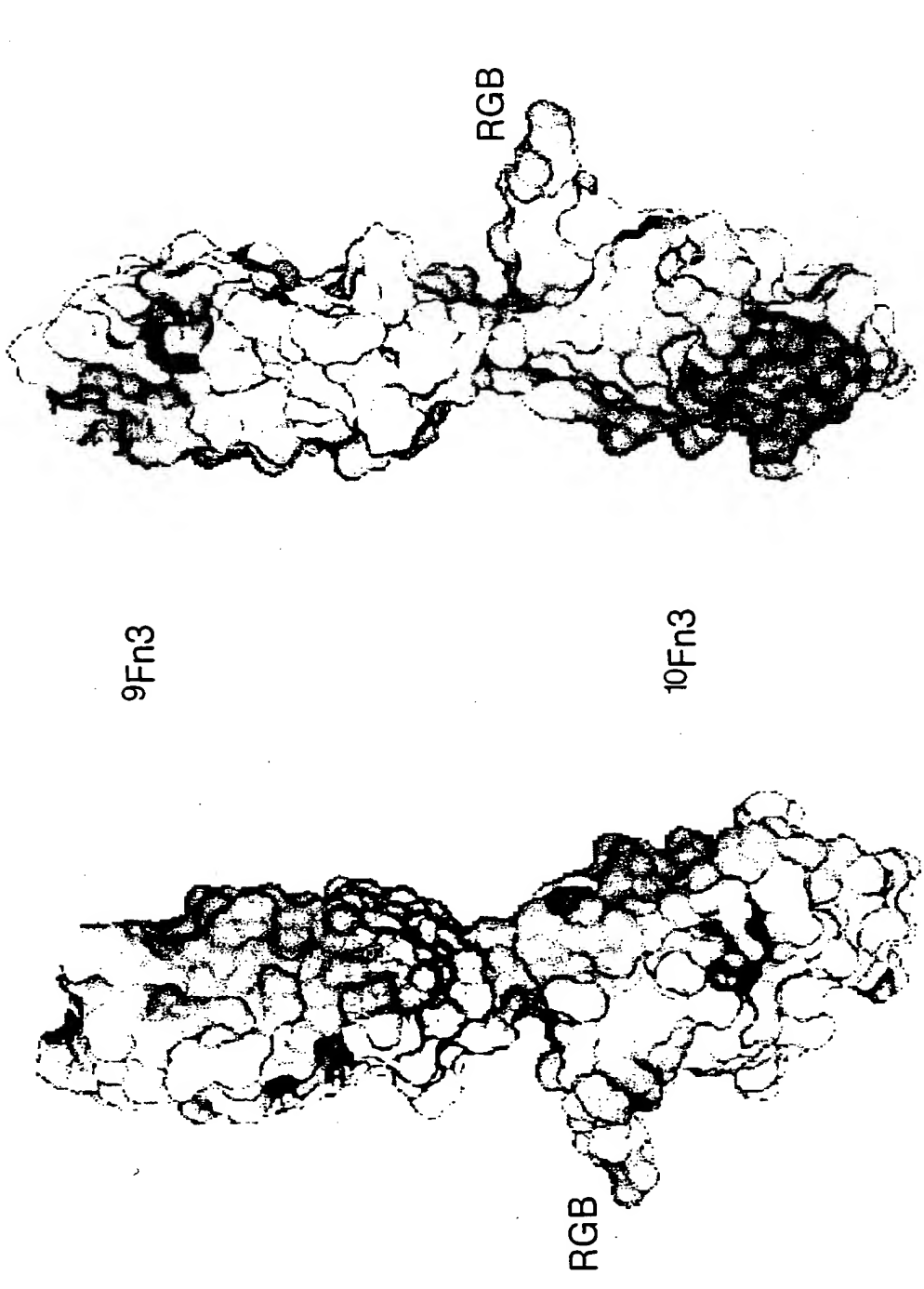



Fig. 6

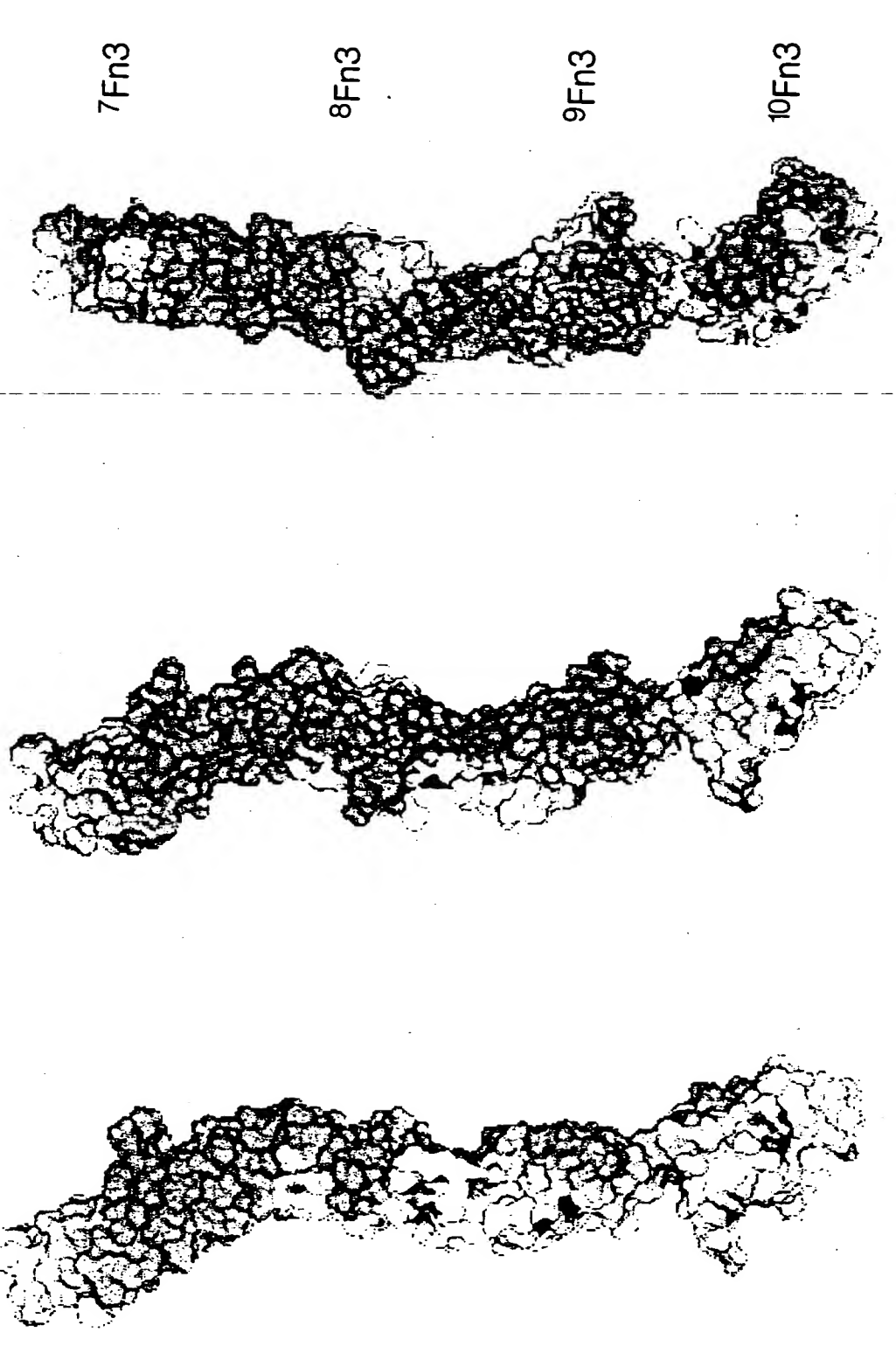



Fig. 7

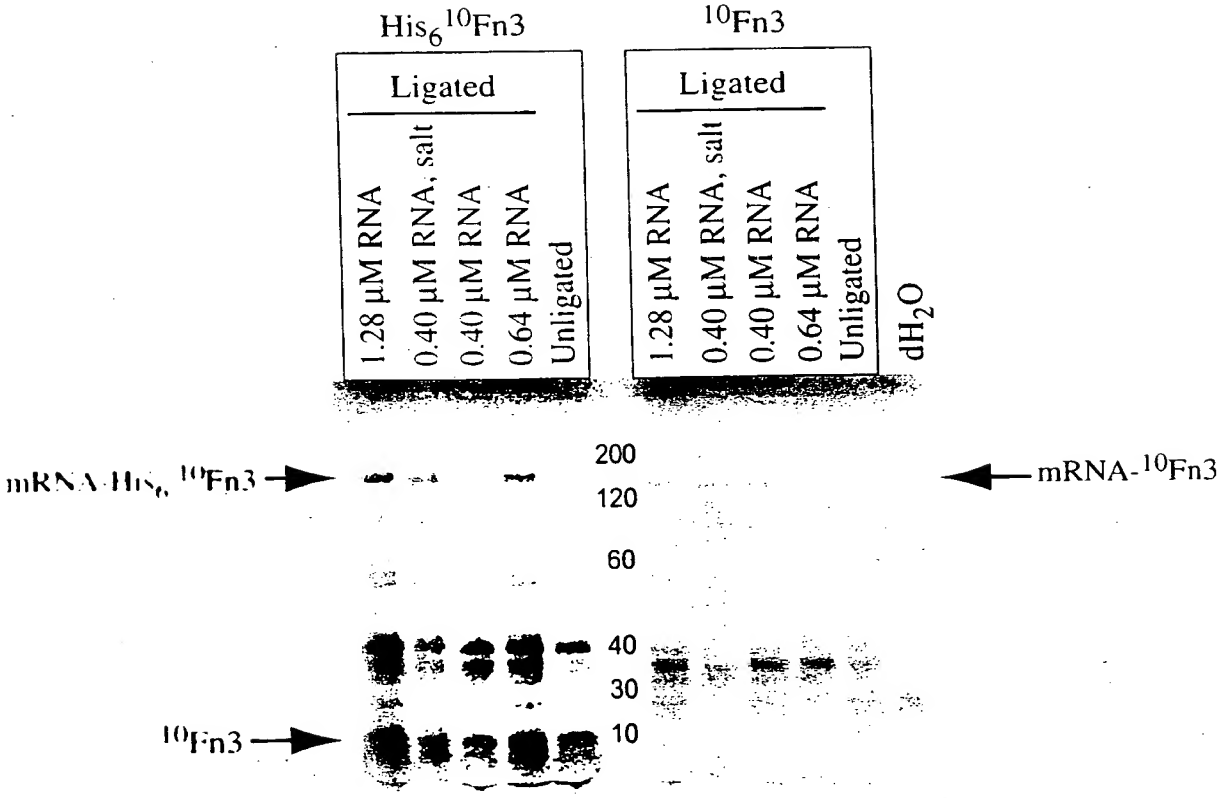

His6 10Fn3
10Fn3
Ligated
Ligated
1.28 µM RNA
0.40 µM RNA, salt
0.40 µM RNA
0.64 µM RNA
Unligated
1.28 µM RNA
0.40 µM RNA, salt
0.40 µM RNA
0.64 µM RNA
Unligated
dH2O
mRNA-His6 10Fn3
mRNA-10Fn3
200
120
60
40
30
10
10Fn3

Fig. 8

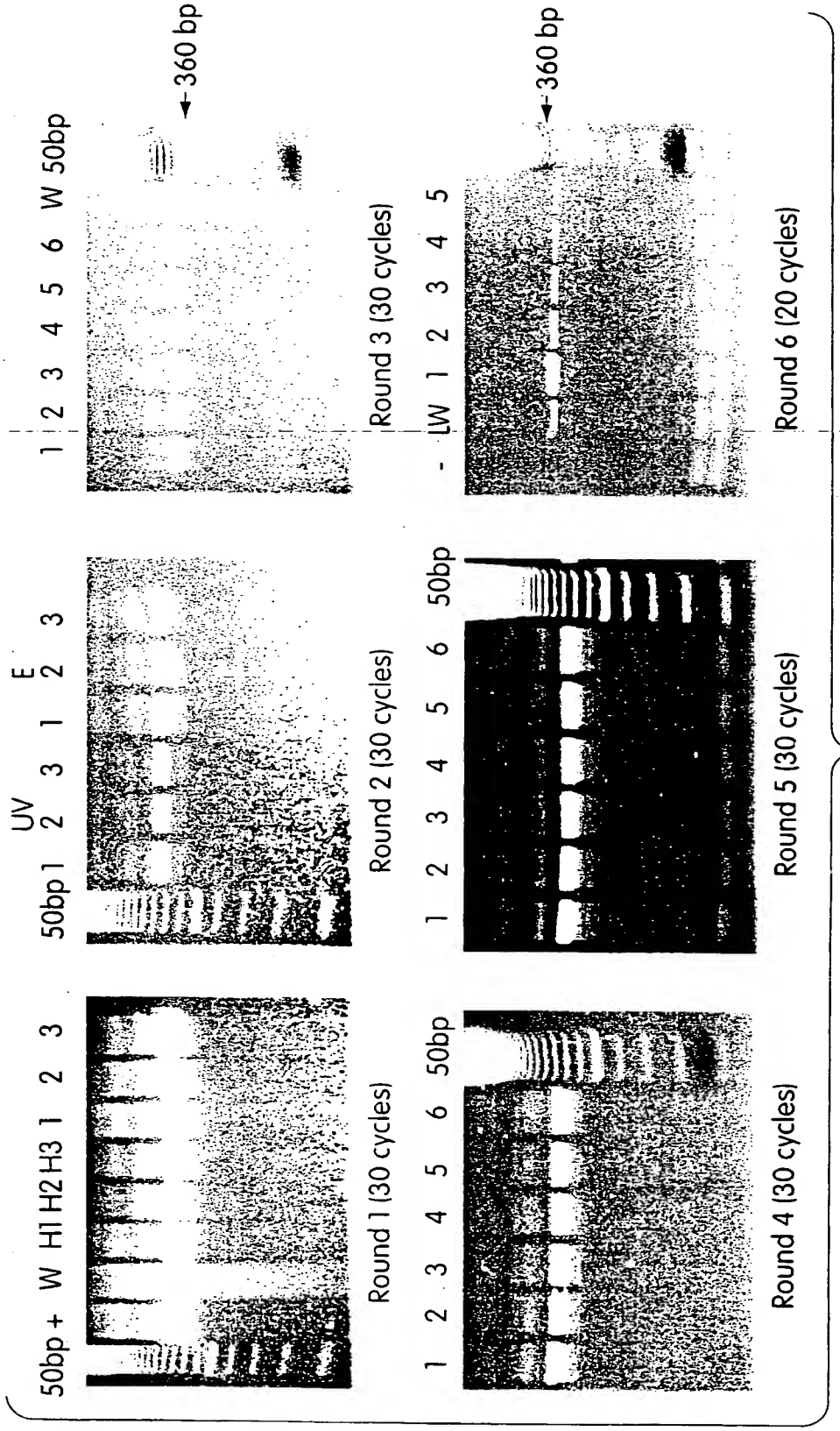

50bp + W H1 H2 H3 1 2 3
Round 1 (30 cycles)
50bp 1 2 3 UV 1 2 3 E
Round 2 (30 cycles)
1 2 3 4 5 6 W 50bp
- 360 bp
Round 3 (30 cycles)
1 2 3 4 5 6 50bp
Round 4 (30 cycles)
1 2 3 4 5 6 50bp
Round 5 (30 cycles)
- LW 1 2 3 4 5
- 360 bp
Round 6 (20 cycles)

Fig. 9

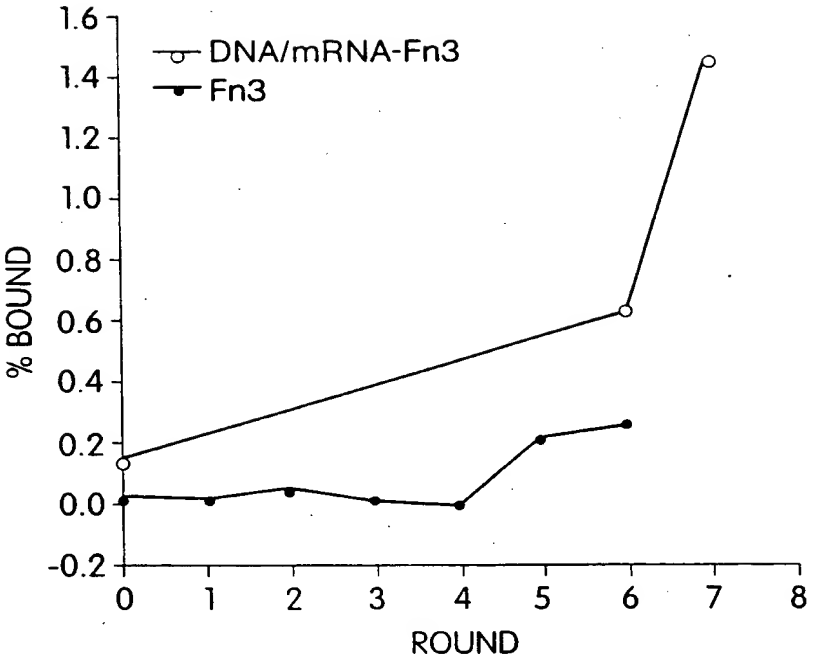



Fig. 10

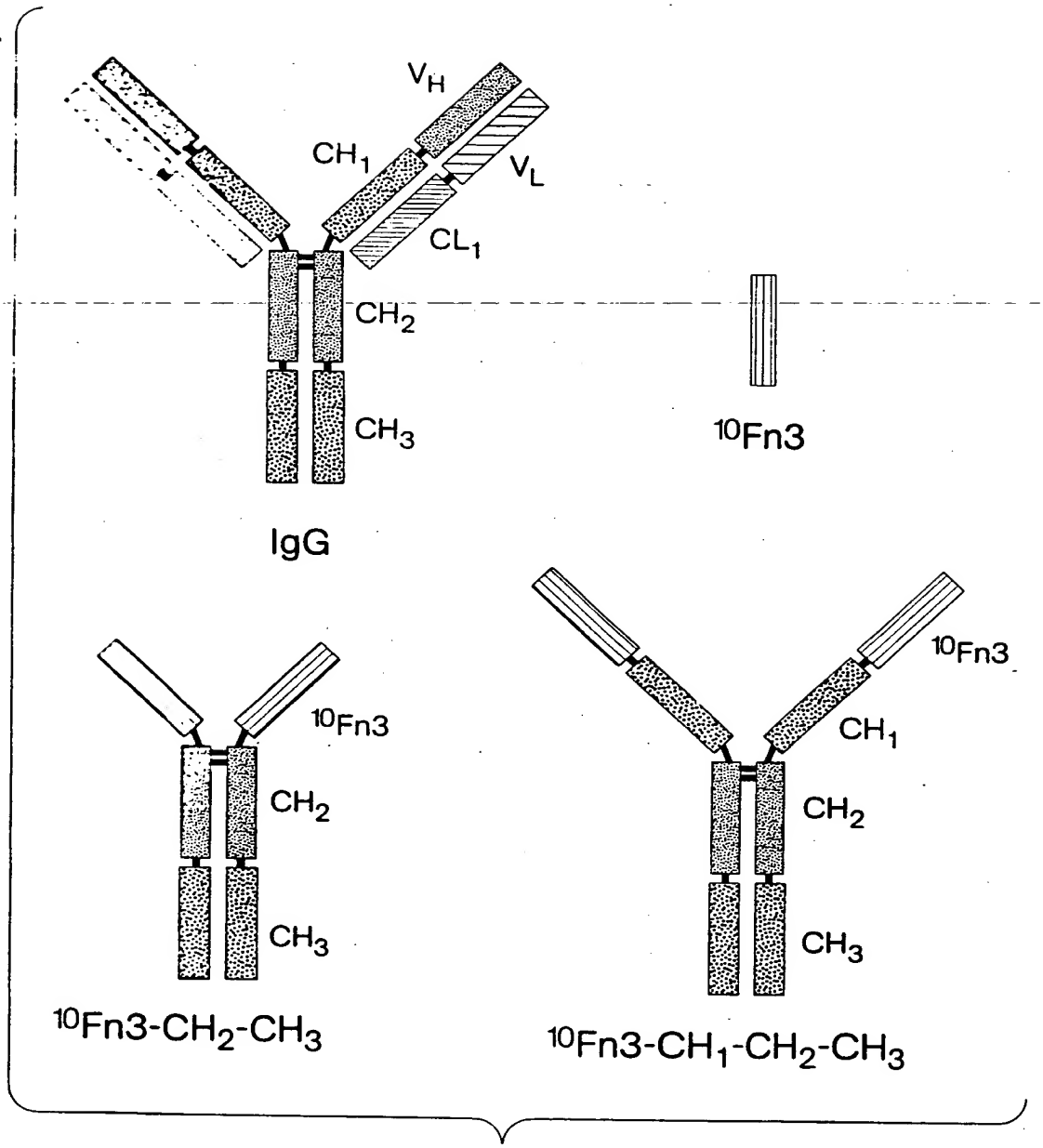

$V_H$
$CH_1$
$V_L$
$CL_1$
$CH_2$
$CH_3$
IgG
$^{10}Fn3$
$^{10}Fn3$
$CH_2$
$CH_3$
$^{10}Fn3$-$CH_2$-$CH_3$
$^{10}Fn3$
$CH_1$
$CH_2$
$CH_3$
$^{10}Fn3$-$CH_1$-$CH_2$-$CH_3$

Fig. 11

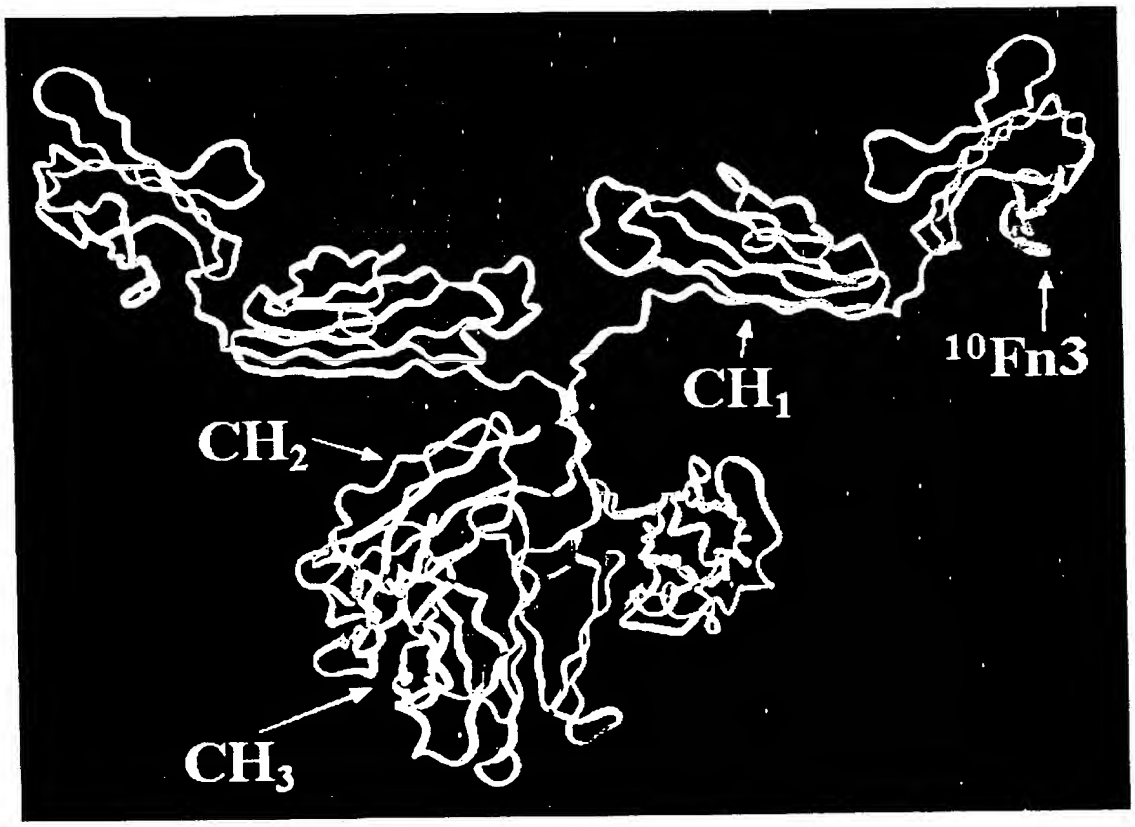



Fig. 12

SEQUENCE LISTING

<110> Phylos, Inc.

<120> PROTEIN SCAFFOLDS FOR ANTIBODY MIMICS
AND OTHER BINDING PROTEINS

<130> 50036/021WO2

<150> 60/111,737
<151> 1998-12-10

<160> 21

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 122
<212> DNA
<213> Homo sapiens

<400> 1
ggaattccta atacgactca ctatagggac aattactatt tacaattaca atgcatcacc 60
atcaccatca cgtttctgat gttccgaggg acctggaagt tgttgctgcg acccccacca 120
gc 122

<210> 2
<211> 104
<212> DNA
<213> Homo sapiens

<400> 2
ggaattccta atacgactca ctatagggac aattactatt tacaattaca atggtttctg 60
atgttccgag ggacctggaa gttgttgctg cgacccccac cagc 104

<210> 3
<211> 126
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (1)...(126)
<223> n = A,T,C or G

<221> misc_feature
<222> (1)...(126)
<223> s = C or G

<400> 3
agcggatgcc ttgtcgtcgt cgtccttgta gtcgctcttc cctgtttctc cgtaagtgat 60
cctgtaatat ctsnnsnnsn nsnnsnnsnn snnccagctg atcagtaggc tggtgggggt 120
cgcagc 126

<210> 4
<211> 62
<212> DNA
<213> Homo sapiens

<400> 4
agcggatgcc ttgtcgtcgt cgtccttgta gtcgctcttc cctgtttctc cgtaagtgat 60
cc 62

<210> 5
<211> 99
<212> DNA
<213> Homo sapiens

<400> 5
ggaattccta tacgactca ctatagggac aattactatt tacaattaca atgcatcacc 60
atcaccatca cctcttcaca ggaggaaata gccctgtcc 99

<210> 6
<211> 132
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (1)...(132)
<223> n = A,T,C or G

<221> misc_feature
<222> (1)...(132)
<223> s = C or G

<400> 6
agcggatgcc ttgtcgtcgt cgtccttgta gtcgctcttc gtataatcaa ctccaggttt 60
aaggccgctg atggtagctg tsnnsnnsnn snnaggcaca gtgaactcct ggacagggct 120
atttcctcct gt 132

<210> 7
<211> 64
<212> DNA
<213> Homo sapiens

<400> 7
agcggatgcc ttgtcgtcgt cgtccttgta gtcgctcttc gtataatcaa ctccaggttt 60
aagg 64

<210> 8
<211> 101
<212> DNA
<213> Homo sapiens

<400> 8
ggaattccta tacgactca ctatagggac aattactatt tacaattaca atgcatcacc 60
atcaccatca cctcttctat accatcactg tgtatgctgt c 101

<210> 9
<211> 114
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (1)...(114)
<223> n = A,T,C or G

<221> misc_feature
<222> (1)...(114)
<223> s = C or G

<400> 9
agcggatgcc ttgtcgtcgt cgtccttgta gtctgttcgg taattaatgg aaattggsnn 60
snnsnnsnns nnsnnsnnsn nsnnsnnagt gacagcatac acagtgatgg tata 114

<210> 10
<211> 57
<212> DNA
<213> Homo sapiens

<400> 10
agcggatgcc ttgtcgtcgt cgtccttgta gtctgttcgg taattaatgg aaattgg 57

<210> 11
<211> 45
<212> DNA
<213> T7 phage and tobacco mosaic virus

<400> 11
gcgtaatacg actcactata gggacaatta ctatttacaa ttaca 45

<210> 12
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> Flag sequence

<400> 12
agcggatgcc ttgtcgtcgt cgtccttgta gtc 33

<210> 13
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Splint oligonucleotide

<221> misc_feature
<222> (1)...(19)

<223> n = A,T,C or G

<400> 13
ttttttttn agcggatgc 19

<210> 14
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Puromycin linker oligonucleotide

<400> 14
aaaaaaaaaa aaaaaaaacc 20

<210> 15
<211> 30
<212> DNA
<213> Mus musculus

<400> 15
gcggcagggt ttgcttactg gggccaaggg 30

<210> 16
<211> 27
<212> DNA
<213> Mus musculus

<400> 16
gggaggggtg gaggtaggtc acagtcc 27

<210> 17
<211> 30
<212> DNA
<213> Mus musculus

<400> 17
tttgctagct ttaccaggag agtgggaggc 30

<210> 18
<211> 33
<212> DNA
<213> Mus musculus

<400> 18
aaaaagcttg ccaaaacgac acccccatct gtc 33

<210> 19
<211> 33
<212> DNA
<213> Mus musculus

<400> 19
catatggttt ctgatattcc gagagatctg gag 33

<210> 20
<211> 43
<212> DNA
<213> Mus musculus

<400> 20
catatgcatc accatcacca tcacgtttct gatattccga gag 43

<210> 21
<211> 30
<212> DNA
<213> Mus musculus

<400> 21
gaattcctat gttttataat tgatggaaac 30

# INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/29317

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :G01N 33/536; C07K 14/00; C12N 15/00
US CL :435/7.1, 7.6, 69.1; 436/86, 87, 536; 530/350

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/7.1, 7.6, 69.1; 436/86, 87, 536; 530/350

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| A | MARKLAND et al. Iterative Optimization of High-Affinity Protease Inhibitors Using Phage Display. 1. Plasmin. Biochemistry. 1996, Vol. 35, No. 24, pages 8045-8057. | 1-52 |
| A | NORD et al. Binding Proteins Selected from Combinatorial Libraries of an α-helical bacterial receptor domain. Nature Biotechnology. August 1997, Vol. 15, pages 772-777. | 1-52 |
| A | KU et al. Alternate Protein Framework for Molecular Recognition. Proc. Natl. Acad. Sci. USA. July 1995, Vol. 92, pages 6552-6556. | 1-52 |

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

| | | | |
|---|---|---|---|
| * | Special categories of cited documents: | "T" | later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention |
| "A" | document defining the general state of the art which is not considered to be of particular relevance | "X" | document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone |
| "E" | earlier document published on or after the international filing date | "Y" | document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art |
| "L" | document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) | "&" | document member of the same patent family |
| "O" | document referring to an oral disclosure, use, exhibition or other means | | |
| "P" | document published prior to the international filing date but later than the priority date claimed | | |

Date of the actual completion of the international search
13 MARCH 2000

Date of mailing of the international search report
06 APR 2000

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703) 305-3230

Authorized officer
HOLLY SCHNIZER
Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet) (July 1998)*

# INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/29317

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

STN (medline, biosis, embase, caplus), EAST (uspat, derwent, EPO), search terms: fibronectin, fn3, type III domain, randomized, antibody mimics, scaffolds, directed evolution, complementary-determining region, integrin binding, minibody, tendamistat.